

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 14:53:43 ; Search time 17 Seconds
(without alignments)
857.546 Million cell updates/sec

Title: US-09-831-805A-6

Perfect score: 310

Sequence: 1 MALRRPPLRLCARLPDFL.....VNYINTDEGFRHKSSFVI 310

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	8	2.6	80	1 TX3A PHONI
2	8	2.6	342	1 ICXX_HUMAN
3	8	2.6	366	1 IC02 GORGO
4	8	2.6	366	1 IC03 GORGO
5	8	2.6	366	1 IC04 GORGO
6	8	2.6	366	1 IC11_HUMAN
7	8	2.6	435	1 YGK1_SALTY
8	7	2.3	161	1 TATB_STRCO
9	7	2.3	224	1 XLRI_MOUSE
10	7	2.3	244	1 TNFC_HUMAN
11	7	2.3	302	1 HTPX_AQUAE
12	7	2.3	306	1 TNFC_MOUSE
13	7	2.3	310	1 TNFC_HARMO
14	7	2.3	333	1 GPR8_HUMAN
15	7	2.3	343	1 HMD_METVO
16	7	2.3	374	1 TRM1_SULTO
17	7	2.3	403	1 TVRP_ECOLI
18	7	2.3	417	1 PVR_CERA
19	7	2.3	417	1 PVR_HUMAN
20	7	2.3	454	1 APY_SOLTU
21	7	2.3	481	1 LMRA_STRLN
22	7	2.3	491	1 CPB4_RABIT
23	7	2.3	491	1 CPB5_RABIT
24	7	2.3	524	1 GTR2_HUMAN
25	7	2.3	574	1 IRL2_MOUSE
26	7	2.3	595	1 IF2P_ARCFU
27	7	2.3	617	1 PYS1_PSEAE
28	7	2.3	643	1 S212_HUMAN
29	7	2.3	643	1 S212_MOUSE
30	7	2.3	643	1 S212_RAT
31	7	2.3	688	1 PYS2_PSEAE
32	7	2.3	693	1 NCPR_ASPNG
33	7	2.3	697	1 TGM2_CHICK

RESULT 1

ALIGNMENTS

34	7	2.3	721	1	YU1Y_ECOLI
35	7	2.3	758	1	SC16_YEAST
36	7	2.3	840	1	CC16_YEAST
37	7	2.3	909	1	RPS2_ARATH
38	7	2.3	981	1	SCA4_RICE
39	7	2.3	987	1	K6P1_CANAL
40	7	2.3	991	1	SCA4_RICSI
41	7	2.3	1011	1	SCA4_RICAF
42	7	2.3	1011	1	SCA4_RICWN
43	7	2.3	1012	1	SCA4_RICSL
44	7	2.3	1013	1	SCA4_RICRH
45	7	2.3	1018	1	SCA4_RICJA
46	7	2.3	1022	1	SCA4_RICCN
47	7	2.3	1053	1	ITA3_MOUSE
48	7	2.3	1066	1	ITA3_CRISP
49	7	2.3	1066	1	ITA3_HUMAN
50	7	2.3	1105	1	YSGE_ECOLI
51	7	2.3	1126	1	HFS5_MOUSE
52	7	2.3	1142	1	PAK1_YEAST
53	7	2.3	1169	1	EX5B_BORBU
54	7	2.3	1234	1	YNK5_CAEEL
55	7	2.3	1234	1	PLE1_HUMAN
56	7	2.3	5065	1	EPPL_HUMAN
57	6	1.9	38	1	CRS3_NOTGO
58	6	1.9	46	1	LHA1_ECTHA
59	6	1.9	57	1	ANDP_DROWA
60	6	1.9	57	1	ANDP_DROME
61	6	1.9	57	1	ANDP_DROSE
62	6	1.9	57	1	Y160_ARCFU
63	6	1.9	60	1	ANDP_DROSI
64	6	1.9	61	1	RS14_MYCQA
65	6	1.9	64	1	YDT6_SCHPO
66	6	1.9	69	1	BR1P_RANPI
67	6	1.9	74	1	SRP_SOYBN
68	6	1.9	76	1	SECG_BACSU
69	6	1.9	82	1	TX32_PHONI
70	6	1.9	83	1	TX31_PHONI
71	6	1.9	86	1	PTHP_STRMU
72	6	1.9	87	1	PTHP_BACHD
73	6	1.9	87	1	PTHP_STREO
74	6	1.9	87	1	PTHP_STRESL
75	6	1.9	94	1	YE23_HAEIN
76	6	1.9	96	1	FER1_AQUAE
77	6	1.9	99	1	Y010_BPPI1
78	6	1.9	102	1	KCRB_PIG
79	6	1.9	109	1	V42_BPT3
80	6	1.9	115	1	NRG4_MOUSE
81	6	1.9	115	1	NU3M_DUGDU
82	6	1.9	115	1	TKN1_RABIT
83	6	1.9	118	1	YKD7_YEAST
84	6	1.9	119	1	B2MG_HUMAN
85	6	1.9	119	1	B2MG_MACFA
86	6	1.9	119	1	B2MG_PONPY
87	6	1.9	123	1	Y055_MYCOE
88	6	1.9	124	1	SPEH_THETN
89	6	1.9	125	1	FRDD_MYCTU
90	6	1.9	130	1	DHSC_PARDE
91	6	1.9	130	1	KV5G_MOUSE
92	6	1.9	130	1	Y060_BPT4
93	6	1.9	130	1	YFFQ_ECOLI
94	6	1.9	134	1	YFID_BACSU
95	6	1.9	138	1	NIKR_ARCFU
96	6	1.9	143	1	G10_ORYSA
97	6	1.9	147	1	YD12_MYCTU
98	6	1.9	148	1	PFID2_ARATH
99	6	1.9	148	1	VE6_HPV34
100	6	1.9	151	1	YG98_CLOAB

P39396 escherichia
P18759 saccharomyc
P09798 saccharomyc
Q42484 arabidopsis
Q9a137 rickettsia
Q94201 candida alb
Q9a177 rickettsia
Q9a183 rickettsia
Q9a182 rickettsia
Q9a180 rickettsia
Q9a181 rickettsia
Q9a179 rickettsia
Q52658 rickettsia
Q62470 mus musculu
P17852 cricetidae
P26006 homo sapien
P38097 escherichia
P59438 mus musculu
P38990 saccharomyc
O51578 borrelia bu
P34578 caenorhabdi
O15149 homo sapien
P58107 homo sapien
P15534 nototadarus
P80100 ectothiorho
O16825 drosophila
P21663 drosophila
Q8wv2 drosophila
Q28419 archaeglob
Q8wsv4 drosophila
O52345 mycoplasma
O14211 schizosacch
Q8qf95 rana pipien
Q07502 glycine max
O32233 bacillus su
O76201 phoneutria
O76200 phoneutria
P45596 streptococc
Q9wvk8 streptococc
P24366 streptococc
P44194 haemophilus
O67065 aquifex aeo
P51712 bacterioph
Q29594 sus scrofa
P20316 bacterioph
Q9wvx4 mus musculu
Q8w9m9 dugong dugo
P41540 oryctolagus
P32858 saccharomyc
P01884 homo sapien
Q8epw0 macaca faec
P18213 pongo pygma
P47301 mycoplasma
Q8ra74 thermoanaer
O10763 mycobacteri
Q59659 paracoccus
P01639 mus musculu
P39222 bacterioph
P76548 escherichia
P54720 bacillus su
O29515 archaeglob
P35682 oryza sativ
Q10620 mycobacteri
Q91398 arabidopsis
P36811 human papil
P33661 clostridium

```
TX3A_PHONI
ID TX3A_PHONI STANDARD; PRT; 80 AA.
AC P81793;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurotoxin PN3A precursor.
OS Phoneutria nigriventer (Brazilian armed spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Entelegynae; Lycosoidea; Ctenidae; Phoneutria.
OX NCBI_TaxID=6918;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland; PubMed=9839681;
EX MEDLINE=99053416;
RA Kalapothakis E., Penaforte C.L., Leao R.M., Cruz J.S., Prado V.F.,
RA Cordeliro M.N., Diniz C.R., Romano-Silva M.A., Prado M.A.M.,
RA Gomez M.V., Beirao P.S.L.;
RT "Cloning, cDNA sequence analysis and patch clamp studies of a toxin
RT from the venom of the armed spider (Phoneutria nigriventer).";
RL Toxicon 36:1971-1980(1998)
CC -!- FUNCTION: Antagonist of L-type calcium channels (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Expressed by the venom gland.
CC -!- SIMILARITY: BELONGS TO THE SPIDER TOXIN TX3 FAMILY.
DR InterPro; IPR004169; spidertoxin.
DR Pfam; PF02819; spidertoxin; 1.
KW Calcium channel inhibitor; Toxin; Neurotoxin; Ionic channel inhibitor;
KW Signal.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 37 BY SIMILARITY.
FT CHAIN 38 71 NEUROTOXIN PN3A.
FT PROPEP 72 80 BY SIMILARITY.
SQ SEQUENCE 80 AA; 8937 MW; 55BF209257EB6793 CRC64;

Query Match 2.6%; Score 8; DB 1; Length 80;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 LALITLGI 263
DB 12 LALITLGI 19

RESULT 2
ID ICXX_HUMAN STANDARD; PRT; 342 AA.
AC P10321;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE HLA class I histocompatibility antigen, C-4 alpha chain.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=86033791; PubMed=3863816;
RA Davidson W.F., Kress M., Khoury G., Jay G.;
RT "Comparison of HLA class I gene sequences. Derivation of
RT locus-specific oligonucleotide probes specific for HLA-A, HLA-B, and
RT HLA-C genes.";
RL J. Biol. Chem. 260:13414-13423(1985).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X60251; CAA42803.1; -.
```

```
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M11886; AAA52665.1; -.
PIR; A24512; HLHUC4.
HSSP; P30685; 1A9E.
MIW; 142840; -.
InterPro; IPR007110; Ig-like.
InterPro; IPR003597; Ig_cl.
InterPro; IPR003006; Ig_MHC.
InterPro; IPR001039; MHC_I.
Pfam; PF00047; Ig; 1.
Pfam; PF00129; MHC_I; 1.
PRINTS; PR01638; MHCCLASSI.
PRODOM; PD000050; MHC_I; 1.
SMART; SMO0407; IGcl; 1.
PROSITE; PS00835; IG_LIKE; 1.
PROSITE; PS00290; IG_MHC; FALSE NEG.
MHC_I; Transmembrane; Glycoprotein.
DOMAIN 25 90
DOMAIN 91 182 EXTRACELLULAR ALPHA-1.
DOMAIN 183 274 EXTRACELLULAR ALPHA-2.
DOMAIN 275 284 EXTRACELLULAR ALPHA-3.
TRANSMEM 285 309 CONNECTING PEPTIDE.
DOMAIN 310 342
DOMAIN 343 386 CYTOPLASMIC TAIL.
CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
DISULFID 164 203 BY SIMILARITY.
SQ SEQUENCE 342 AA; 38082 MW; 22C39A6D84C05D09 CRC64;

Query Match 2.6%; Score 8; DB 1; Length 342;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 VLVVLAVL 256
DB 293 VLVVLAVL 300

RESULT 3
ID IC02_GORGO STANDARD; PRT; 366 AA.
AC P30385;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Class I histocompatibility antigen, GOGO-C0201 alpha chain precursor.
OS Gorilla gorilla gorilla (Lowland gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OX NCBI_TaxID=9595;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=92078860; PubMed=1744581;
RA Lawlor D.A., Warren E., Taylor P., Parham P.;
RT "Gorilla class I major histocompatibility complex alleles: comparison
RT to human and chimpanzee class I.";
RL J. Exp. Med. 174:1491-1509(1991).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X60251; CAA42803.1; -.
```

```

DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001039; MHC_I.
DR Pfam; PF00047; Ig_1.
DR Pfam; PF00129; MHC_I_1.
DR PRINTS; PR01638; MHCCLASSI.
DR PRODOM; PD000050; MHC_I_1.
DR SMART; SM00407; Igcl; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; FALSE NEG.
DR KEGG; K04491; Transmembrane; Glycoprotein; Signal.
KW MHC_I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 366
FT FT
FT FT
FT DOMAIN 25 114
FT DOMAIN 115 206
FT DOMAIN 207 298
FT DOMAIN 299 308
FT TRANSMEM 309 332
FT DOMAIN 333 366
FT DISULFID 125 188
FT DISULFID 227 283
FT CARBOHYD 110 110
FT SEQUENCE 366 AA; 40782 MW; 11CCADIF6091831B CRC64;
SQ
Query Match 2.6%; Score 8; DB 1; Length 366;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 249 VLVWLAVL 256
DB 317 VLVWLAVL 324
|||||
|||||

RESULT 5
1C04_GORGO
IID 1C04_GORGO STANDARD; PRT; 366 AA.
AC P30387;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Class 1 histocompatibility antigen, GGO-C0203 alpha chain precursor.
OS Gorilla gorilla gorilla (Lowland gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OC NCBI_TaxID=9595;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92078860; PubMed=1744581;
RZ Lawlor D.A., Warren E., Taylor P., Parham P.;
RT "Gorilla class I major histocompatibility complex alleles: comparison
RT to human and chimpanzee class I.";
RT J. Exp. Med. 174:1491-1509(1991).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X60248; CAA42800.1; --
CC PIR; JH0547; JH0547.
CC HSP; P30685; 1A9F.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003597; Ig_c1.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR001039; MHC_I.

```

```

DR Pfam; PF00047; ig; 1.
DR Pfam; PF00129; MHC_I; 1.
DR PRINTS; PR01638; MHCCLASSI.
DR ProDom; PD000050; MHC_I; 1.
DR SMART; SM00407; IGcl; 1.
DR PROSITE; PS00835; IG LIKE; 1.
DR PROSITE; PS00290; IG_MHC; FALSE NEG.
KW MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 366
FT DOMAIN 25 114
FT DOMAIN 115 206
FT DOMAIN 207 298
FT DOMAIN 299 308
FT DOMAIN 309 332
FT TRANSMEM 333 366
FT DISULFID 125 188
FT DISULFID 227 283
FT CARBOHYD 110 110
SQ SEQUENCE 366 AA; 40970 MW; EE962C8189CAC001 CRC64;

Query Match 2.6%; Score 8; DB 1; Length 366;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 VLWVLAVL 256
DB 317 VLWVLAVL 324

RESULT 6
ICIL HUMAN
ID ICIL HUMAN STANDARD; PRT; 366 AA.
AC Q29631;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE HLA class I histocompatibility antigen, CW-7 CW*0704 alpha chain
DE precursor.
GN Homo sapiens (Human).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96086482; PubMed=7482492;
RA Vilches C., Bunce M., de Pablo R., Herrero M.J., Kreisler M.;
RT "Anchored PCR cloning of the novel HLA-Cw*0704 allele detected by
RT PCR-SSP.";
RL Tissue Antigens 46:19-23 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Blood;
RC MEDLINE=96232973; PubMed=8655361;
RA Wang H., Tokunaga K., Ishikawa Y., Asahina A., Kuwata S.,
RA Akaza T., Tadokoro K., Shibata Y., Takiguchi M., Juji T.;
RT "Identification and DNA typing of two Cw7 alleles (Cw*0702 and
RT Cw*0704) in Japanese, with the corrected sequence of Cw*0702.";
RL Hum. Immunol. 45:52-58 (1996).
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
CC THE IMMUNE SYSTEM.
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

CC -----
DR EMBL; X83394; CAA58313.1; -.
DR EMBL; D49552; BAA08500.1; -.
DR PIR; I37078; I37078.
DR HSSP; P30460; IAGD.
DR MIM; 142840; -.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG cl.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR001039; MHC_I.
DR Pfam; PF00047; ig; 1.
DR Pfam; PF00129; MHC_I; 1.
DR PRINTS; PR01638; MHCCLASSI.
DR ProDom; PD000050; MHC_I; 1.
DR SMART; SM00407; IGcl; 1.
DR PROSITE; PS00835; IG LIKE; 1.
DR PROSITE; PS00290; IG_MHC; FALSE NEG.
KW MHC I; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 24
FT CHAIN 25 366
FT DOMAIN 25 114
FT DOMAIN 115 206
FT DOMAIN 207 298
FT DOMAIN 299 308
FT TRANSMEM 309 332
FT DOMAIN 333 366
FT DISULFID 125 188
FT DISULFID 227 283
FT CARBOHYD 110 110
SQ SEQUENCE 366 AA; 40819 MW; 4073C3BC53E55D0C CRC64;

Query Match 2.6%; Score 8; DB 1; Length 366;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 249 VLWVLAVL 256
DB 317 VLWVLAVL 324

RESULT 7
YGIK SALTY
ID YGIK SALTY STANDARD; PRT; 435 AA.
AC P40800;
DT 01-FEB-1995 (Rel. 31, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ygiK.
GN YGIK OR STM3171.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RA Cong J., Schmid M.B.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856 (2001).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Probable).
```

CC -!- SIMILARITY: BELONGS TO THE YIAN/YGIK FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U09309; AA056680.1; -;
DR EMBL; AE088845; AAL22045.1; -;
DR StyGene; SG10523; YGIK.
DR InterPro; IPR000252; Deda.
DR InterPro; IPR004681; TRAP_transptDctM.
DR Pfam; PF00597; Deda; 1.
DR TIGRFAMs; TIGR00786; dcmM; 1.
KW Hypothetical protein; Transmembrane; Inner membrane;
FT Complete proteome.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 57 77 POTENTIAL.
FT TRANSMEM 110 130 POTENTIAL.
FT TRANSMEM 146 166 POTENTIAL.
FT TRANSMEM 176 196 POTENTIAL.
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 280 300 POTENTIAL.
FT TRANSMEM 321 341 POTENTIAL.
FT TRANSMEM 343 363 POTENTIAL.
FT TRANSMEM 367 387 POTENTIAL.
FT TRANSMEM 408 428 POTENTIAL.
FT CONFLICT 119 119 A -> R (IN REF. 1).
FT CONFLICT 124 124 A -> P (IN REF. 1).
SQ SEQUENCE 435 AA; 46004 MW; EEA79BFCFE8A58A CRC64;

Query Match 2.6%; Score 8; DB 1; Length 435;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 242 IGGIIGGV 249
|||||
DB 236 IGGIIGGV 243

RESULT 8
TATB_STRCO
ID TATB_STRCO STANDARD; PRT; 161 AA.
AC Q9FBK8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sec-independent protein translocase protein tatB homolog.
GN TATB OR SC05150 OR SCP8.13.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -!- FUNCTION: Required for correct localization of precursor proteins

CC bearing signal peptides with the twin arginine conserved motif
CC S/T-R-X-F-L-K. This sec-independent pathway is termed TAT for
CC twin-arginine translocation system. This system mainly transports
CC proteins with bound cofactors that require folding prior to export
CC (by similarity).
CC -!- SUBCELLULAR LOCATION: Membrane bound (Probable).
CC -!- SIMILARITY: Belongs to the tatB family.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AL939122; CAC01351.1; -;
DR HAWAP; MF_00237; -; 1.
DR InterPro; IPR003998; TatB.
DR PRINTS; PR01506; TATBPROTEIN.
KW Transport; Protein transport; Translocation; Transmembrane;
KW Membrane; Complete proteome.
FT TRANSMEM 2 22 POTENTIAL.
SQ SEQUENCE 161 AA; 17757 MW; 0510C0D221BEEF3 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 161;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 250 LVVLAVL 256
|||||
DB 13 LVVLAVL 19

RESULT 9

XLRI_MOUSE
ID XLRI_MOUSE STANDARD; PRT; 224 AA.
AC Q921L4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE X-linked juvenile retinoschisis protein precursor.
GN RSI OR RSIH OR XLRS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ;
RX MEDLINE=95160474; PubMed=10051329;
RA Gehrig A.E., Warneke-Wittstock R., Sauer C.G., Weber B.H.F.;
RT "Isolation and characterization of the murine X-linked juvenile
RT retinoschisis (Relh) gene";
RL Mamm. Genome 10:303-307(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Eye;
RA Brunner B., Todt T., Lenzner S., Stout K., Schulz U., Ropers H.-H.,
RA Kalscheuer V.M.;
RT "Genomic structure and comparative analysis of seven contiguous genes
RT disclose a large region with conserved gene order in human Xp22.2-
RT p22.1";
RL Submitted (DSC-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=99148018; PubMed=10023077;
RA Reid S.N., Ahmedov N.B., Piriev N.I., Kozak C.A., Danciger M.,
RA Farber D.B.;
RT "The mouse X-linked juvenile retinoschisis cDNA: expression in
RT photoreceptors";
RL Gene 227:257-266(1999).


```
DR InterPro; IPR003636; TNF_subf.
DR Pfam; PF00229; TNF; 1.
DR PRINTS; PRO1234; TNCRGSIISFCT.
DR ProDom; PD002012; TNF_subf; 1.
DR SMART; SM00207; TNF; 1.
DR PROSITE; PS00251; TNF_1; 1.
DR PROSITE; PS0049; TNF_2; 1.
KW Cytokine; Transmembrane; Glycoprotein; Signal-anchor;
KW Alternative splicing; Polymorphism.
FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 19 48 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 49 222 EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 53 77 GLVETADPGAAQOGLGFKLPPEE -> GLGFRSCQRRSQ
FT FT KOISAPGSOLPTS (in isoform 2).
FT FT /FTID=VSP_006441.
FT FT Missing (in isoform 2).
FT FT /FTID=VSP_006442.
FT FT G -> E.
FT FT /FTID=VAR_013025.
FT FT A -> P.
FT FT /FTID=VAR_013026.
FT FT DFGAQAQQGL -> GLSAPGSORT (IN REF. 2;
FT FT AAB37342).
SQ SEQUENCE 244 AA; 25390 MW; F41569459830ED4C CRC64;

Query Match 2.3%; Score 7; DB 1; Length 244;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VLAVLAL 258
| | | | |
DB 40 VLAVLAL 46

RESULT 11
HTPX AQUAE
ID HTPX AQUAE STANDARD; PRT; 302 AA.
AC O67798;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable protease htpx homolog (EC 3.4.24.-).
GN HTPX OR AQ.1991.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=95373320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
CC -!- CORFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use. By non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000766; AAC07747.1; -.
CC PIR; B70471; B70471.
CC MEROPS; M48.004; -.

DR InterPro; IPR001915; Peptidase M48.
DR InterPro; IPR006025; Zn_Mtpeptdse.
DR Pfam; PF01435; Peptidase M48; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW Transmembrane; Hydrolase; Metalloprotease; Zinc; Complete proteome.
FT TRANSMEM 27 47 POTENTIAL.
FT TRANSMEM 151 171 POTENTIAL.
FT TRANSMEM 195 215 POTENTIAL.
FT METAL 141 141 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 142 142 BY SIMILARITY.
FT METAL 145 145 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 302 AA; 33030 MW; 3278BF48156BE895 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 302;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 IGGIIGG 248
| | | | |
DB 31 IGGIIGG 37

RESULT 12
TNFC MOUSE
ID TNFC MOUSE STANDARD; PRT; 306 AA.
AC P41155;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lymphotoxin-beta (LT-beta) (Tumor necrosis factor C) (TNF-C) (Tumor
DE necrosis factor ligand superfamily member 3).
GN LTB OR TNFSF3 OR TNFC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=95148600; PubMed=7846035;
RA Pokholok D.K., Maroulakou I.G., Kuprash D.V., Alimzhanov M.B.,
RA Kozlov S.V., Novobrantseva T.I., Turetskaya R.L., Green J.E.,
RA Nedospasov S.A.;
RT "Cloning and expression analysis of the murine lymphotoxin beta
RT gene."
RL Proc. Natl. Acad. Sci. U.S.A. 92:674-678(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Spleen;
RX MEDLINE=95088371; PubMed=7995944;
RA Lawton P., Nelson J., Tizard R., Browning J.L.;
RT "Characterization of the mouse lymphotoxin-beta gene."
RN [3]
RP SEQUENCE FROM N.A.
RA Rowen L., Qin S., Madan A., Abbasi N., James R., Dickhoff R.,
RA Shaffer I., Ratcliffe A., Loretz C., Laskey S., Hood L.;
RT "Sequence of the mouse major histocompatibility class III region."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Cytokine that binds to LTB/TNFRSF3. May play a specific
CC role in immune response regulation. Provides the membrane anchor
CC for the attachment of the heterotrimeric complex to the cell
CC surface.
CC -!- SUBUNIT: Heterotrimer of either two LTB and one LTA subunits or
CC (less prevalent) two LTA and one LTB subunits (By similarity).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the tumor necrosis factor family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----

DR EMBL; U12029; AAA67716.1; -
 DR EMBL; U16984; AAB60493.1; -
 DR EMBL; U16985; AAA70089.1; -
 DR EMBL; U06950; AAA18592.1; -
 DR EMBL; AF109719; AACB2483.1; -
 DR PIR; I49139; I49139;
 DR MGI; MGI:104796; LCB.
 DR GO; GO:0007515; P:lymph gland development; IMP.
 DR InterPro; IPR006053; TNF abc.
 DR InterPro; IPR006052; TNF_family.
 DR InterPro; IPR003636; TNF_subf.
 DR Pfam; PF00229; TNF; 1.
 DR PRINTS; PR01234; TNECROSISFCT.
 DR ProDom; PD002012; TNF_subf; 1.
 DR SMART; SM00207; TNF; 1.
 DR PROSITE; PS00251; TNF 1; 1.
 DR PROSITE; PS0049; TNF 2; 1.
 DR Cytokine; Transmembrane; Glycoprotein; Signal-anchor.
 KW CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 1 27
 FT TRANSMEM 28 48
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 49 306
 FT CARBOHYD 98 98
 FT CARBOHYD 284 284
 FT N-LINKED (GLCNAC...) (POTENTIAL).
 FT N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 306 AA; 32328 MW; B7D276AB8422549 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 306;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VLAVLAL 258
 DB 40 VLAVLAL 46

RESULT 13

TNFC MARMO
 ID TNFC MARMO STANDARD; PRT; 310 AA.
 AC Q9JW10; Q9JW11;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Lymphotoxin-beta (Lr-beta) (tumor necrosis factor C) (TNF-C) (Tumor
 DE necrosis factor ligand superfamily member 3).
 GN LTB OR TNFSF3 OR TNFC.
 OS Marmota monax (Woodchuck).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Scuridae; Sciurinae;
 OC Marmota.
 OC NCBI_TaxID=9995;
 RP SEQUENCE FROM N.A.
 RN [1]
 RX MEDLINE=20184748; PubMed=10721723;
 RA Li D.H., Havell E.A., Brown C.L., Cullen J.M.;
 RT "Woodchuck lymphotoxin-alpha, -beta and tumor necrosis factor genes:
 RT structure, characterization and biological activity.";
 RL Gene 242:295-305(2000).
 CC -!- FUNCTION: Cytokine that binds to LTBR/TNFRSF3. May play a specific
 CC role in immune response regulation. Provides the membrane anchor
 CC for the attachment of the heterotrimeric complex to the cell
 CC surface.
 CC -!- SUBUNIT: Heterotrimer of either two LTB and one LTA subunits or
 CC (less prevalent) two LTA and one LTB subunits (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
 CC -!- SIMILARITY: Belongs to the tumor necrosis factor family.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----

DR EMBL; AF096268; AAF34866.1; -
 DR EMBL; AF095587; AAF34865.1; -
 DR HSSP; P01374; 1TNR.
 DR InterPro; IPR006053; TNF abc.
 DR InterPro; IPR006052; TNF_family.
 DR InterPro; IPR003636; TNF_subf.
 DR Pfam; PF00229; TNF; 1.
 DR PRINTS; PR01234; TNECROSISFCT.
 DR ProDom; PD002012; TNF_subf; 1.
 DR SMART; SM00207; TNF; 1.
 DR PROSITE; PS0049; TNF 2; 1.
 DR Cytokine; Transmembrane; Glycoprotein; Signal-anchor.
 KW CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 1 27
 FT TRANSMEM 28 48
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 49 310
 FT CARBOHYD 272 272
 FT CARBOHYD 280 280
 FT N-LINKED (GLCNAC...) (POTENTIAL).
 FT N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 310 AA; 32644 MW; 73B354EFC8B3B3BE CRC64;

Query Match 2.3%; Score 7; DB 1; Length 310;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 VLAVLAL 258
 DB 40 VLAVLAL 46

RESULT 14

GPR8 HUMAN
 ID GPR8 HUMAN STANDARD; PRT; 333 AA.
 AC P48146; O9H4K3;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable G protein-coupled receptor GPR8.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96070436; PubMed=7590751;
 RA O'Dowd B.F., Scheideler M.A., Nguyen T., Cheng R., Rasmussen J.S.,
 RA Marchese A., Zastawny R., Heng H.H.Q., Tsui L.-C., Shi X., Asa S.,
 RA Puy L., George S.R.;
 RT "The cloning and chromosomal mapping of two novel human opioid-
 RT somatostatin-like receptor genes, GPR7 and GPR8, expressed in
 RT discrete areas of the brain.";
 RL Genomics 28:84-91(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguely C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Clee C.M.,
 RA Chapman J.C., Clamp M., Clark A.G., Clark L.N., Clark S.Y., Conn N.R.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhali P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,

OS Sulfolobus tokodaii.
OC Archaea: Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_taxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankaï A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Katayama N., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguni A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermophilic
RT Crenarchaeon, Sulfolobus tokodaii strain 7";
RL DNA Res. 8:123-140(2001).
CC -!- FUNCTION: Dimethylates a single guanine residue at position 26 of
CC a number of tRNAs using S-adenosyl-L-methionine as donor of the
CC methyl groups (By similarity).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
CC homocysteine + tRNA containing N(2)-methylguanine.
CC -!- SIMILARITY: BELONGS TO THE TRM1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP000985; BAB6311.1; -;
DR HAMAP; MF_00290; -; 1.
DR InterPro; IPR002905; TRM.
DR Pfam; PF02005; TRM; 1.
KW Transferase; Methyltransferase; tRNA processing; Complete proteome.
SQ SEQUENCE 374 AA; 43333 MW; 0B4E6E3C2420B15F CRC64;

Query Match 2.3%; Score 7; DB 1; Length 374;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 124 KEIDEIV 130
DB 362 KEIDEIV 368

RESULT 17
TYRP_ECOLI
ID TYRP_ECOLI STANDARD; PRT; 403 AA.
AC P18199; P76309;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Tyrosine-specific transport protein (Tyrosine permease).
GN TYRP OR B1307 OR SF1953.
OS Escherichia coli, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 623;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli;
RX MEDLINE=89008121; PubMed=3049553;
RA Wooley P.J., Pittard A.J.;
RT "DNA sequence of the gene (tyrp) encoding the tyrosine-specific
RT transport system of Escherichia coli.";
RL J. Bacteriol. 170:4946-4949(1988).
RN [2]
RP REVISIONS.
RC SPECIES=E.coli;

RX MEDLINE=91216998; PubMed=2022620;
RA Howitt S.;
RL Unpublished results, cited by;
RL Sarsero J.P., Wooley P.J., Gollnick P., Yanofsky C., Pittard A.J.;
RL J. Bacteriol. 173:3231-3234(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayaashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sempel G., Seki Y., Sivasubraman S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RL corresponding to the 40.1-50.0 min region on the linkage map.";
RN DNA Res. 3:379-392(1996).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX MEDLINE=2227406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang X., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
CC -!- FUNCTION: INVOLVED IN TRANSPORTING TYROSINE ACROSS THE CYTOPLASMIC
CC MEMBRANE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -!- INDUCTION: REPRESSED BY TYROSINE AND INDUCED BY PHENYLALANINE
CC UNDER THE CONTROL OF REGULATORY PROTEIN TYRR.
CC -!- SIMILARITY: Belongs to the amino acid/polyamine transporter family
CC II. Mcr / tnaB / tyrp permease subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M23240; AAA24705.1; -;
DR EMBL; M23240; AAA24705.1; -;
DR EMBL; D90832; BAA15730.1; -;
DR EMBL; AE015214; AAN43504.1; -;
DR PIR; C64954; GREY.
DR EcoGene; EG11041; tyrp.
DR InterPro; IPR002293; AA/rel_permease1.
DR InterPro; IPR002422; AA/rel_permease2.
DR InterPro; IPR002091; AAA_permease.
DR Pfam; PF0322; Trrp_Tyr_perm; 1.
DR PRINTS; PR00166; AROAPRMEASE.
DR TIGRfams; TIGR00837; ataaP; 1.
DR PROSITE; PS00594; AROMATIC_AA_PERMEASE_1; 1.
KW Transport; Amino-acid transport; Transmembrane; Inner membrane;
KW Complete proteome.
RN 9
RP REVISIONS.
RC TRANSMEM 35
FT POTENTIAL.
FT 55

```

FT TRANSMEM 81 101 POTENTIAL.
FT TRANSMEM 122 142 POTENTIAL.
FT TRANSMEM 148 168 POTENTIAL.
FT TRANSMEM 184 204 POTENTIAL.
FT TRANSMEM 217 237 POTENTIAL.
FT TRANSMEM 275 295 POTENTIAL.
FT TRANSMEM 308 328 POTENTIAL.
FT TRANSMEM 332 352 POTENTIAL.
FT TRANSMEM 376 396 POTENTIAL.
FT TRANSMEM 131 135 AGGVV -> RRVAVL (IN REF. 1).
FT CONFLICT 222 224 SAI -> ECD (IN REF. 1).
SQ SEQUENCE 403 AA; 42819 MW; 7BECCAA833679796 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 403;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 253 LAVIALI 259
Db 344 LAVIALI 350

RESULT 18
PVR_CERAE STANDARD; PRT; 417 AA.
AC P32506;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Poliovirus receptor precursor.
GN PVR OR PVS.
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Cercopithecus.
OX NCBI_TaxID=95334;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS ALPHA AND DELTA).
RC TISSUE=Kidney;
RX MEDLINE=93059651; PubMed=1331508;
RA Koike S., Ise I., Sato Y., Yonekawa H., Gotoh O., Nomoto A.;
RT "A second gene for the African green monkey poliovirus receptor that
RT has no putative N-glycosylation site in the functional N-terminal
RT immunoglobulin-like domain.";
RL J. Virol. 66:7059-7066(1992).
CC -!- FUNCTION: NOT KNOWN. USED BY POLIOVIRUS TO BIND AND ENTER THE
CC CELL.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (ALPHA FORM);
CC SECRETED (BETA AND GAMMA FORMS).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=Alpha;
CC IsoId=P32506-1; Sequence=Displayed;
CC Name=Beta;
CC IsoId=P32506-3; Sequence=Not described;
CC Name=Gamma;
CC IsoId=P32506-4; Sequence=Not described;
CC Name=Delta;
CC IsoId=P32506-2; Sequence=VSP_002622, VSP_002623;
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D12611; BAA02136.1; -
CC EMBL; D12612; BAA02137.1; -

```

```

DR PIR; A44194; A44194.
DR PIR; B44194; B44194.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig; 3.
DR SMART; SMO0406; Igv; 1.
DR PROSITE; PS00835; IG_LIKE; 3.
KW Immunoglobulin domain; Receptor; Transmembrane; Glycoprotein; Signal;
KW Repeat; Alternative splicing.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 417 POLIOVIRUS RECEPTOR.
FT DOMAIN 21 343 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 344 367 POTENTIAL.
FT DOMAIN 368 417 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 24 139 IG-LIKE V-TYPE.
FT DOMAIN 145 237 IG-LIKE C2-TYPE 1.
FT DOMAIN 244 328 IG-LIKE C2-TYPE 2.
FT DISULFID 49 123 BY SIMILARITY.
FT DISULFID 166 221 BY SIMILARITY.
FT DISULFID 266 312 BY SIMILARITY.
FT CARBOHYD 85 85 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 120 120 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 237 237 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 307 307 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 313 313 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 386 392 EHASASA -> HHQSCHN (in isoform Delta).
FT VARSPLIC 393 417 Missing (in isoform Delta).
FT VARSPLIC 417 AA; 45464 MW; DA4AD0FE4D2F6E1F CRC64;
SQ SEQUENCE 417 AA; 45464 MW; DA4AD0FE4D2F6E1F CRC64;

Query Match 2.3%; Score 7; DB 1; Length 417;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 132 ELTVQVK 138
Db 324 ELTVQVK 330

RESULT 19
PVR_HUMAN STANDARD; PRT; 417 AA.
AC P15151; P15152; Q15267; Q15268;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Poliovirus receptor precursor (CD155 antigen).
GN PVR OR PVS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89168426; PubMed=2538245;
RA Mendelsohn C.L., Wimmer E., Racaniello V.R.;
RT "Cellular receptor for poliovirus: molecular cloning, nucleotide
RT sequence, and expression of a new member of the immunoglobulin
RT superfamily.";
RL Cell 56:855-865 (1989).
RN [2]
RN REVISIONS.
RA Racaniello V.R.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=91006015; PubMed=2170108;
RA Koike S., Horie H., Ise I., Okitsu A., Yoshida M., Iizuka N.,
RA Takeuchi K., Takegami T., Nomoto A.;

```


DE Apyrase precursor (EC 3.6.1.5) (ATP-diphosphatase) (Adenosine
DE diphosphatase) (ADPase) (ATP-diphosphohydrolase).
GN RROPI.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
OC Asteridae; lamiales; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 59-160; 236-253 AND 332-345.
RC TISSUE=tuber;
RX MEDLINE=96158985; PubMed=8579614;
RA Handa M., Guidotti G.;
RT "Purification and cloning of a soluble ATP-diphosphohydrolase
RT (apyrase) from potato tubers (Solanum tuberosum).";
RL Biochem. Biophys. Res. Commun. 218:916-923(1996).
RN [2]
RP SEQUENCE OF 42-54; 68-95 AND 236-253.
RC STRAIN=cv. Desiree;
RX MEDLINE=96355615; PubMed=8703025;
RA Vasconcelos E.G., Ferreira S.T., de Carvalho T.M.U., de Souza W.,
RA Kettlun A.M., Mancilla M., Valenzuela M.A., Verjovski-Almeida S.;
RT "Partial purification and immunohistochemical localization of ATP
RT diphosphohydrolase from Schistosoma mansoni. Immunological cross-
RT reactivities with potato apyrase and Toxoplasma gondii nucleoside
RT triphosphate hydrolase.";
RL J. Biol. Chem. 271:22139-22145(1996).
CC [1]- FUNCTION: CATALYZES THE HYDROLYSIS OF PHOSPHOANHYDRIDE BONDS OF
CC NUCLEOSIDE TRI- AND DI-PHOSPHATES.
CC [2]- CATALYTIC ACTIVITY: ATP + 2 H(2)O = AMP + 2 phosphate.
CC [3]- SUBCELLULAR LOCATION: Membrane-associated (Probable).
CC [4]- PTM: THE N-TERMINUS IS BLOCKED.
CC [5]- SIMILARITY: BELONGS TO THE GDAL / CD39 NTPASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U58597; AAB02720.1; -.
CC PIR; JC4616; JC4616.
DR InterPro; IPR000407; GDAL_CD39_NTPase.
DR Pfam; PF01150; GDAL_CD39_1.
DR PROSITE; PS01238; GDAL_CD39_NTPASE; 1.
KW Hydrolase; Transmembrane; Calcium; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 454 APYRASE.
FT TRANSMEM 426 446 POTENTIAL.
FT CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 454 AA; 50041 MW; 9D9FE431DA2F52F CRC64;
Query Match 2.3%; Score 7; DB 1; Length 454;
Best Local Similarity 100.0%; Pred. No. 59; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 88 LAGRAEI 94
DB 250 LAGRAEI 256
RESULT 21
ID_LMRA_STRLN STANDARD; PRT; 481 AA.
AC P46104;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Lincomycin resistance protein.

GN LMRA.
OS Streptomyces lincolnensis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1915;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=78-11;
RX MEDLINE=93023841; PubMed=1328813;
RA Zhang H.Z., Schmidt H., Piepersberg W.;
RT "Molecular cloning and characterization of two lincomycin-resistance
RT genes, lmra and lmrb, from Streptomyces lincolnensis 78-11.";
RL Mol. Microbiol. 6:2147-2157(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=78-11;
RX MEDLINE=96020646; PubMed=8577249;
RA Peschke U., Schmidt H., Zhang H.Z., Piepersberg W.;
RT "Molecular characterization of the lincomycin-production gene cluster
RT of Streptomyces lincolnensis 78-11.";
RL Mol. Microbiol. 16:1137-1156(1995).
CC [1]- FUNCTION: PROTON-DEPENDENT TRANSPORTER. MAY MEDIATE THE EFFLUX OF
CC LINCOMYCIN.
CC [2]- SUBCELLULAR LOCATION: Integral membrane protein.
CC [3]- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC AS THE DRUG RESISTANCE TRANSLUCASE FAMILY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X59926; CAA42550.1; -.
CC EMBL; X79146; CAA55745.1; -.
DR PIR; S69808; S69808.
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub_transporter.
DR InterPro; IPR005829; Sug_transporter.
DR InterPro; IPR001411; TCR_TetB.
DR Pfam; PF00083; sugar_tr; 1.
DR PRINTS; PR01036; TCR_TetB.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN 1.
KW Transport; Transmembrane; Antibiotic resistance.
FT TRANSMEM 30 50 POTENTIAL.
FT TRANSMEM 67 87 POTENTIAL.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 127 147 POTENTIAL.
FT TRANSMEM 162 182 POTENTIAL.
FT TRANSMEM 185 205 POTENTIAL.
FT TRANSMEM 215 235 POTENTIAL.
FT TRANSMEM 245 265 POTENTIAL.
FT TRANSMEM 285 305 POTENTIAL.
FT TRANSMEM 318 338 POTENTIAL.
FT TRANSMEM 340 360 POTENTIAL.
FT TRANSMEM 374 394 POTENTIAL.
FT TRANSMEM 421 441 POTENTIAL.
FT TRANSMEM 446 466 POTENTIAL.
SQ SEQUENCE 481 AA; 50421 MW; DBF494C51D7B6F17 CRC64;
Query Match 2.3%; Score 7; DB 1; Length 481;
Best Local Similarity 100.0%; Pred. No. 62; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 252 VLAVLAL 258
DB 461 VLAVLAL 467
RESULT 22
CPB4_RABIT

ID CPB4 RABBIT STANDARD; PRT; 491 AA.
AC P00178; P00177;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome P450 2B4 (EC 1.14.14.1) (CYP2B4) (P450-LM2) (Isozyme 2)
DE (P450 types B0 and B1).
GN CYP2B4.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88094290; PubMed=2826996;
RA Gasser R., Negishi M., Philpot R.M.;
RT "Primary structures of multiple forms of cytochrome P-450 isozyme 2
derived from rabbit pulmonary and hepatic cDNAs.";
RL Mol. Pharmacol. 33:22-30(1988).
RN [1]
RN [2]
RP SEQUENCE.
RX MEDLINE=84042509; PubMed=6579541;
RA Tarr G.E., Black S.D., Fujica V.S., Coon M.J.;
RT "Complete amino acid sequence and predicted membrane topology of
phenobarbital-induced cytochrome P-450 (isozyme 2) from rabbit liver
microsomes.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:6552-6556(1983).
RN [3]
RP SEQUENCE.
RX MEDLINE=83160983; PubMed=6833251;
RA Heinemann F.S., Ozols J.;
RT "The complete amino acid sequence of rabbit phenobarbital-induced
liver microsomal cytochrome P-450";
RL J. Biol. Chem. 258:4195-4201(1983).
RN [4]
RP MUTAGENESIS OF CYS-436.
RX MEDLINE=2222627; PubMed=12237221;
RA Vatis K.P., Peng H.-M., Coon M.J.;
RT Replacement of active-site cysteine-436 by serine converts cytochrome
P450 2B4 into an NADPH oxidase with negligible monooxygenase
activity.";
RL J. Inorg. Biochem. 91:542-553(2002).
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN
NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY
OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY
ACIDS, AND XENOBIOTICS. IN THE EPOXIDATION OF ARACHIDONIC ACID
IT HAS A UNIQUE PREFERENCE FOR THE 5,6-OLEFIN.
CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
oxidized flavoprotein + H(2)O.
CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -!- INDUCTION: By phenobarbital.
CC -!- POLYMORPHISM: TYPES B0 AND B1 ARE PROBABLY ALLELIC VARIANTS.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; M20856; AAA65840.1; -
CC EMBL; M20857; AAA31224.1; -
CC PIR; A00179; O4RBFC.
CC HSSP; P00179; 1D76.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450.1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Iron;
Heme; Microsome; Endoplasmic reticulum; Phosphorylation; Polymorphism.

FT MOD RES 128 128 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
FT METAL 436 436 IRON (HEME AXIAL LIGAND).
FT VARIANT 39 39 V -> I (IN B1).
FT VARIANT 174 174 I -> V (IN B1).
FT VARIANT 290 290 L -> I (IN B1).
FT VARIANT 314 314 M -> L (IN B1).
FT VARIANT 420 420 L -> M (IN B1).
FT CONFLICT 91 91 Q -> E (IN REF. 3).
FT CONFLICT 95 95 FS -> SF (IN REF. 3).
FT CONFLICT 99 100 MISSING (IN REF. 3).
FT CONFLICT 135 136 EG -> GY (IN REF. 3).
FT CONFLICT 193 193 P -> K (IN REF. 3).
FT CONFLICT 221 221 P -> S (IN REF. 3).
FT CONFLICT 303 303 T -> A (IN REF. 3).
FT CONFLICT 461 465 SPVP -> GNLSL (IN REF. 3).
SQ SEQUENCE 491 AA; 55713 MW; 0DB943C6DCDF690B CRC64;
Query Match 2.3%; Score 7; DB 1; Length 491;
Best Local Similarity 100.0%; Pred.No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 20 LLLFRG 26
DB 16 LLLFRG 22
RESULT 23
ID CPB5 RABBIT STANDARD; PRT; 491 AA.
AC P12789;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome P450 2B5 (EC 1.14.14.1) (CYP2B5) (P450 type B2) (P450 form
HPI).
GN CYP2B5.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88094290; PubMed=2826996;
RA Gasser R., Negishi M., Philpot R.M.;
RT "Primary structures of multiple forms of cytochrome P-450 isozyme 2
derived from rabbit pulmonary and hepatic cDNAs.";
RL Mol. Pharmacol. 33:22-30(1988).
RN [2]
RP SEQUENCE OF 1-56 FROM N.A.
RX MEDLINE=95383381; PubMed=7654758;
RA Lehner M., Schulze J., Petzold A., Bernhardt R., Hlavica P.;
RT "Rabbit liver cytochrome P-450 2B5: high-level expression of the
full-length protein in Escherichia coli, purification, and catalytic
activity.";
RL Biochim. Biophys. Acta 1245:107-115(1995).
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN
NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY
OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY
ACIDS, AND XENOBIOTICS.
CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
oxidized flavoprotein + H(2)O.
CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum.
CC -!- INDUCTION: By phenobarbital.
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC

```

CC -----
DR EMBL; M20855; AAA31223.1; -
DR EMBL; S78830; AAB35177.1; -
DR PIR; S31278; S31278.
DR HSSP; P00179; 1D76.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
KW Microsome; Endoplasmic reticulum; Phosphorylation.
FT MOD_RES 128 128 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
FT METAL 436 436 IRON (HEME AXIAL LIGAND).
SQ SEQUENCE 491 AA; 55769 MW; ABE2B978B8408DCFCRC64;
Query Match 2.3%; Score 7; DB 1; Length 491;
Best Local Similarity 100.0%; Pred.No.63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 20 LLLLFPG 26
Db 16 LLLLFPG 22
RESULT 24
GTR2 HUMAN
ID GTR2 HUMAN STANDARD; PRT; 524 AA.
AC P11168;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Solute carrier family 2, facilitated glucose transporter, member 2
DE (Glucose transporter type 2, liver).
GN SLC2A2 OR GLUT2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT LEU-68.
RC TISSUE=Kidney, and Liver;
RX MEDLINE=88289735; PubMed=3399500;
RA Fukumoto H., Sieno S., Imura H., Sieno Y., Eddy R.L., Fukushima Y.,
RA Byers M.G., Shows T.B., Bell G.I.;
RT "Sequence, tissue distribution, and chromosomal localization of mRNA
RL encoding a human glucose transporter-like protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:5434-5438(1988).
CC -!- FUNCTION: FACILITATIVE GLUCOSE TRANSPORTER. THIS ISOFORM LIKELY
CC MEDIATES THE BIDIRECTIONAL TRANSFER OF GLUCOSE ACROSS THE PLASMA
CC MEMBRANE OF HEPATOCYTES AND IS RESPONSIBLE FOR UPTAKE OF GLUCOSE
CC BY THE BETA CELLS; MAY COMPRISE PART OF THE GLUCOSE-SENSING
CC MECHANISM OF THE BETA CELL. MAY ALSO PARTICIPATE WITH THE
CC NA(+)/GLUCOSE COTRANSPORTER IN THE TRANSCELLULAR TRANSPORT OF
CC GLUCOSE IN THE SMALL INTESTINE AND KIDNEY.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: LIVER, INSULIN-PRODUCING BETA CELL, SMALL
CC INTESTINE, AND KIDNEY.
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. GLUCOSE
CC TRANSPORTER SUBFAMILY.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03810; AAB59514.1; -
DR PIR; A31318; A31318.
DR Genew; HGNC:11006; SLC2A2.
DR GK; P11168; -.
DR MIM; 138160; -.

```

```

DR GO; GO:0005887; C: integral to plasma membrane; TAS.
DR GO; GO:0005624; C: membrane fraction; TAS.
DR GO; GO:0005355; F: glucose transporter activity; TAS.
DR GO; GO:0005975; P: carbohydrate metabolism; TAS.
DR GO; GO:0015758; P: glucose transport; TAS.
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub_transporter.
DR InterPro; IPR005829; Sug_transporter.
DR InterPro; IPR003663; Sugar_transp.
DR Pfam; PF00083; sugar_tr; 1.
DR PRINTS; PR00171; SUGRTENSPORT.
DR TIGRFAMs; TIGR00879; SF; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Transmembrane; Sugar transport; Transport; Glycoprotein;
KW Multigene family; Polymorphism.
FT DOMAIN 1 10 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 11 31 1 (POTENTIAL).
FT DOMAIN 32 98 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 99 119 2 (POTENTIAL).
FT DOMAIN 120 127 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 128 148 3 (POTENTIAL).
FT DOMAIN 149 158 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 159 179 4 (POTENTIAL).
FT DOMAIN 180 187 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 188 208 5 (POTENTIAL).
FT DOMAIN 209 217 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 218 238 6 (POTENTIAL).
FT DOMAIN 239 303 7 (POTENTIAL).
FT TRANSMEM 304 324 8 (POTENTIAL).
FT DOMAIN 325 338 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 339 359 9 (POTENTIAL).
FT DOMAIN 360 368 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 369 389 10 (POTENTIAL).
FT DOMAIN 390 400 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 401 421 11 (POTENTIAL).
FT DOMAIN 422 433 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 434 454 12 (POTENTIAL).
FT DOMAIN 455 461 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 462 482 13 (POTENTIAL).
FT DOMAIN 483 524 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 62 62 N-LINKED (GLCNAC... ) (POTENTIAL).
FT VARIANT 68 68 P -> L.
FT VARIANT 101 101 /FTID=VAR_007169.
FT VARIANT 110 110 V -> I (IN dbSNP:18000572).
FT VARIANT 110 110 /FTID=VAR_014718.
FT VARIANT 478 478 T -> I (IN dbSNP:5400).
FT VARIANT 478 478 /FTID=VAR_014719.
FT VARIANT 478 478 L -> V (IN dbSNP:5397).
SQ SEQUENCE 524 AA; 57489 MW; DA60057207EC083 CRC64;
Query Match 2.3%; Score 7; DB 1; Length 524;
Best Local Similarity 100.0%; Pred.No.66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 201 TGTIVFT 207
Db 7 TGTIVFT 13
RESULT 25
IRL2_MOUSE
ID IRL2_MOUSE STANDARD; PRT; 574 AA.
AC Q9RSR7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Interleukin 1 receptor-like 2 precursor (IL-1Rrp2).
GN IL1RL2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

```



```

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20459050; PubMed=10882729;
RA Born T.L.; Smith D.E., Garika K.E., Renshaw B.R., Bertles J.S.,
RT "Identification and characterization of two members of a novel class
RT class of interleukin-1 receptor (IL-1R) family. Delineation of a new
RL J. Biol. Chem. 275:23946-23954(2000).
CC -!- FUNCTION: Receptor for interleukin 1 family member 9 (IL1F9).
CC Binding to the agonist leads to the activation of NF-kappa B (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- MISCELLANEOUS: Does not bind interleukin 1 alpha (IL-1A) or
CC interleukin 1 beta (IL-1B) (By similarity).
CC -!- SIMILARITY: BELONGS TO THE INTERLEUKIN-1 RECEPTOR FAMILY.
CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 TIR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF284433; AAG21367.1; -.
CC HSSP; P14778; 1IRA.
CC MGD; MGI:1913107; Il1rl2.
CC InterPro; IPR007110; IG-like.
CC InterPro; IPR003599; IG.
CC InterPro; IPR003006; IG_MHC.
CC InterPro; IPR004075; IL1_receptor1.
CC InterPro; IPR004074; IL1_receptor1i.
CC InterPro; IPR004076; IL1R receptor.
CC InterPro; IPR000157; TIR_domain.
CC Pfam; PF00047; ig_3.
CC Pfam; PF01582; TIR; 1.
CC SMART; SM00409; IG; 2.
CC SMART; SM00255; TIR; 1.
CC PROSITE; PS50835; IG LIKE; 2.
CC PROSITE; PS50104; TIR; 1.
CC Immunoglobulin domain; Transmembrane; Glycoprotein; Receptor; Signal;
KW Repeat.
FT SIGNAL 1 21
FT CHAIN 22 574
FT DOMAIN 22 338
FT TRANSF 339 359
FT TRANSF 360 574
FT DOMAIN 25 113
FT DOMAIN 132 215
FT DOMAIN 225 321
FT DOMAIN 384 542
FT DISULF 44 97
FT DISULF 149 199
FT DISULF 252 319
FT CARBOHYD 43 43
FT CARBOHYD 55 55
FT CARBOHYD 111 111
FT CARBOHYD 231 231
FT CARBOHYD 227 227
FT CARBOHYD 253 253
FT CARBOHYD 269 269
FT CARBOHYD 290 290
FT CARBOHYD 302 302
FT SEQUENCE 574 AA; 65108 MW; A677A77BBFA50A76 CRC64;
Query Match 2.3%; Score 7; DB 1; Length 574;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 18 FFLLLLF 24
DB 12 FFLLLLF 18
RESULT 26
ID IF2P_ARCFU STANDARD; PRT; 595 AA.
AC 029490;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable translation initiation factor IF-2.
GN INFB OR AF0768.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=989475;
RA Klenk H.-P., Clayton R.A., Tomb J.-P., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervilave A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness B.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -!- FUNCTION: FUNCTION IN GENERAL TRANSLATION INITIATION BY PROMOTING
CC THE BINDING OF THE FORMYL METHIONINE-TRNA TO RIBOSOMES. SEEMS TO
CC FUNCTION ALONG WITH EIF-2 (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001051; AAB90465.1; -.
CC TIGR; AF0768; -.
CC HAWAP; MF_00100; -.
CC InterPro; IPR000795; EF_GTPbind.
CC InterPro; IPR004161; EFTU_D2.
CC InterPro; IPR001178; IF2.
CC InterPro; IPR005225; Small_GTP.
CC InterPro; IPR004544; TIF_aIF-2.
CC Pfam; PF00009; GTP_EFTU; 1.
CC Pfam; PF03144; GTP_EFTU_D2; 1.
CC PRINTS; PR00315; ELONGATNFCT.
CC PRODOM; PD186100; IF2; 1.
CC TIGRFAMS; TIGR00491; aIF-2; 1.
CC TIGRFAMS; TIGR00231; small_GTP; 1.
CC PROSITE; PS01176; IF2; FALSE_NEG.
KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT NP_BIND 20 27 GTP (BY SIMILARITY).
FT NP_BIND 81 85 GTP (BY SIMILARITY).
FT NP_BIND 135 138 GTP (BY SIMILARITY).
SQ SEQUENCE 595 AA; 66680 MW; AFED672ED0D9703 CRC64;
Query Match 2.3%; Score 7; DB 1; Length 595;

```



```
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 294 IRTDEEG 300
Db 353 IRTDEEG 359

RESULT 27
PYS1_PSEAE STANDARD; PRT; 617 AA.
AC Q05583;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pycocin S1 (EC 3.1.1.-) (Killer protein).
GN PYS1.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-12 AND 483-503.
RC STRAIN=NH-H;
RX MEDLINE=93259934; PubMed=8491711;
RA Sano Y., Matsui H., Kobayashi M., Kageyama M.;
RT "Molecular structures and functions of pycocins S1 and S2 in
RT Pseudomonas aeruginosa.";
RL J. Bacteriol. 173:2907-2916(1993).
CC -!- FUNCTION: CAUSES BREAKDOWN OF CHROMOSOMAL DNA AS WELL AS COMPLETE
CC -!- INHIBITION OF LIPID SYNTHESIS IN SENSITIVE CELLS.
CC -!- SUBUNIT: PURIFIED PYCOCIN S1 MAKES UP A COMPLEX OF THE TWO (LARGE
CC AND SMALL) PROTEINS. THE LARGE PROTEIN, BUT NOT THE PYCOCIN
CC COMPLEX, SHOWS IN VITRO DNASE ACTIVITY.
CC -!- MISCELLANEOUS: PYCOCINS CONTAIN N-TERMINAL RECEPTOR-BINDING DOMAIN,
CC TRANSLOCATION DOMAIN AND C-TERMINAL DNASE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE NUCLEASE FAMILY OF COLICINS AND
CC PYOSINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; D12707; BAA02201.1; -.
DR PIR; A36907; A36907.
DR HSSP; Q47112; 7CBI.
DR InterPro; IPR002711; HNH.
DR InterPro; IPR003615; HNH_nuc.
DR InterPro; IPR003060; Pycocin_killer.
DR Pfam; PF01844; HNH; 1.
DR PRINTS; PR01300; PYCOKILLER.
DR SMART; SM00507; HNHc; 1.
KW Antibiotic; Bacteriocin; Hydrolase; Endonuclease.
FT INIT MET 0
FT SEQUENCE 617 AA; 65498 MW; 74C8046EAE9CAF7 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 617;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 146 VPKAVPV 152
Db 380 VPKAVPV 386

RESULT 28
S212_HUMAN STANDARD; PRT; 643 AA.
ID S212_HUMAN
AC Q92959;
```

```
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Solute carrier family 21 member 2 (Prostaglandin transporter) (PGT).
GN SLC21A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96379664; PubMed=8787677;
RA Lu R., Kanai N., Bao Y., Schuster V.L.;
RT "Cloning, in vitro expression, and tissue distribution of a human
RT prostaglandin transporter cDNA(hPGT).";
RL J. Clin. Invest. 98:1142-1149(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98289612; PubMed=9618293;
RA Lu R., Schuster V.L.;
RT "Molecular cloning of the gene for the human prostaglandin transporter
RT hPGT: gene organization, promoter activity, and chromosomal
RT localization.";
RL Biochem. Biophys. Res. Commun. 246:805-812(1998).
CC -!- FUNCTION: May mediate the release of newly synthesized
CC prostaglandins from cells, the transepithelial transport of
CC prostaglandins, and the clearance of prostaglandins from the
CC circulation. Transports PGD2, as well as PGE1, PGE2 and PGF2A.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: BELONGS TO THE SLC21 FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; U70867; AAC09469.1; -.
DR EMBL; AF056732; AAC62004.1; JOINED.
DR EMBL; AF056719; AAC62004.1; JOINED.
DR EMBL; AF056720; AAC62004.1; JOINED.
DR EMBL; AF056721; AAC62004.1; JOINED.
DR EMBL; AF056722; AAC62004.1; JOINED.
DR EMBL; AF056723; AAC62004.1; JOINED.
DR EMBL; AF056724; AAC62004.1; JOINED.
DR EMBL; AF056725; AAC62004.1; JOINED.
DR EMBL; AF056726; AAC62004.1; JOINED.
DR EMBL; AF056727; AAC62004.1; JOINED.
DR EMBL; AF056728; AAC62004.1; JOINED.
DR EMBL; AF056729; AAC62004.1; JOINED.
DR EMBL; AF056730; AAC62004.1; JOINED.
DR EMBL; AF056731; AAC62004.1; JOINED.
DR Genes; HGNC:10955; SLC21A2.
DR MIM; 601460; -.
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0005319; F:lipid transporter activity; TAS.
DR GO; GO:0006869; P:lipid transport; TAS.
DR InterPro; IPR004157; OATP_Cterm.
DR InterPro; IPR004156; OATP_Nterm.
DR Pfam; PF03137; OATP_C; 1.
DR Pfam; PF03132; OATP_N; 1.
DR TIGRFAMs; TIGR00805; oat; 1.
DR Transmembrane; Transport; Glycoprotein.
KW TRANSMEM 27 47
FT TRANSMEM 71 91
FT TRANSMEM 101 121
FT TRANSMEM 172 192
FT TRANSMEM 210 230
FT TRANSMEM 258 278
```

```

FT TRANSMEM 328 348 POTENTIAL.
FT TRANSMEM 366 386 POTENTIAL.
FT TRANSMEM 400 420 POTENTIAL.
FT TRANSMEM 511 531 POTENTIAL.
FT TRANSMEM 551 571 POTENTIAL.
FT TRANSMEM 606 626 POTENTIAL.
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 643 AA; 70116 MW; 7FC434AS3ED62C9E CRC64;

Query Match 2.3%; Score 7; DB 1; Length 643;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 LIGAVNL 34
DB 367 LIGAVNL 373

RESULT 29
S212 MOUSE
ID S212 MOUSE STANDARD; PRT; 643 AA.
AC Q9EPT5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Solute carrier family 21 member 2 (Prostaglandin transporter) (PGT).
GN SLC21A2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND MUTAGENESIS OF VAL-610 AND
RP ILE-611.
RC TISSUE=Lung;
RX MEDLINE=99414062; PubMed=10484490;
RA Pucci M.L., Bao Y., Chan B., Itoh S., Lu R., Copeland N.G.,
RA Gilbert D.J., Jenkins N.A., Schuster V.L.;
RT "Cloning of mouse prostaglandin transporter PGT cDNA: species-specific
RT substrate affinities.";
RL Am. J. Physiol. 277:R734-R741(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC STRAIN=FVB/N; TISSUE=Breast Cancer;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunatathne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: May mediate the release of newly synthesized
CC prostaglandins from cells, the trans epithelial transport of
CC prostaglandins, and the clearance of prostaglandins from the
CC circulation. Transports PG2, as well as PG1, PG2 and PGF2A.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- ALTERNATIVE PRODUCTS:

```

```

CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9EPT5-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9EPT5-2; Sequence=VSP_006128, VSP_006129;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Highly expressed in lung and liver. Detected
CC at lower levels in kidney and skeletal muscle.
CC -!- SIMILARITY: BELONGS TO THE SLC21 FAMILY OF TRANSPORTERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF323958; AAG40332.1; -;
CC EMBL; BC035200; AAH35200.1; -;
CC GGD; MGI:1346021; SLC21a2.
CC GO; GO:0015132; F:prostaglandin transporter activity; IDA.
CC GO; GO:0015732; P:prostaglandin transport; IDA.
CC InterPro; IPR007114; MPS.
CC InterPro; IPR004157; OATP_Cterm.
CC Pfam; PF03137; OATP_C; 1.
CC TIGRFAMs; TIGR00805; oat; 1.
CC Transmembrane; transport; Glycoprotein; Alternative splicing.
KW TRANSMEM 31 51
FT TRANSMEM 72 92 POTENTIAL.
FT TRANSMEM 101 121 POTENTIAL.
FT TRANSMEM 172 192 POTENTIAL.
FT TRANSMEM 210 230 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 321 341 POTENTIAL.
FT TRANSMEM 365 385 POTENTIAL.
FT TRANSMEM 399 419 POTENTIAL.
FT TRANSMEM 510 530 POTENTIAL.
FT TRANSMEM 550 570 POTENTIAL.
FT TRANSMEM 605 625 POTENTIAL.
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 490 490 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 368 414
FT
FT VARSPLIC 415 643
FT
FT MUTAGEN 610 610 V->M: NO EFFECT.
FT MUTAGEN 611 611 I->G: NO EFFECT.
FT CONFLICT 3 3 L -> F (IN REF. 1).
SQ SEQUENCE 643 AA; 70146 MW; AC487386F847D9D8 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 643;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 LIGAVNL 34
DB 366 LIGAVNL 372

RESULT 30
S212 RAT
ID S212 RAT STANDARD; PRT; 643 AA.
AC Q00910;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Solute carrier family 21 member 2 (Prostaglandin transporter) (PGT)

```

DE (Matrin F/G).
GN SLC21A2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=91296785; PubMed=2068100;
RA Hakes D.J., Berzney R.;
RT "Molecular cloning of matrin F/G: A DNA binding protein of the nuclear
matrix that contains putative zinc finger motifs.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:6186-6190(1991).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95273959; PubMed=7754369;
RA Kanai N., Lu R., Satriano J.A., Bao Y., Wolkoff A.W., Schuster V.L.;
RT "Identification and characterization of a prostaglandin transporter.";
RL Science 268:866-869(1995).
CC -!- FUNCTION: May mediate the release of newly synthesized
prostaglandins from cells, the trans epithelial transport of
prostaglandins, and the clearance of prostaglandins from the
circulation. Transports PG2, as well as PGE1, PGE2 and PGF2A.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE SLC21 FAMILY OF TRANSPORTERS.
CC -!- CAUTION: Was originally (Ref.1) thought to be a nuclear DNA-
binding protein.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; M64862; AAA41574.1; ALT_INIT.
DR PIR; A41120; A41120.
DR InterPro; IPR004157; OATP_Cterm.
DR InterPro; IPR004156; OATP_Nterm.
DR Pfam; PF03137; OATP_C; 1.
DR Pfam; PF03132; OATP_N; 1.
DR TIGRFAMs; TIGR00805; oat; 1.
KW Transmembrane; Transport; Glycoprotein.
FT TRANSMEM 27 47
FT TRANSMEM 71 91 POTENTIAL.
FT TRANSMEM 101 121 POTENTIAL.
FT TRANSMEM 172 192 POTENTIAL.
FT TRANSMEM 210 230 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 321 341 POTENTIAL.
FT TRANSMEM 365 385 POTENTIAL.
FT TRANSMEM 399 419 POTENTIAL.
FT TRANSMEM 512 532 POTENTIAL.
FT TRANSMEM 550 570 POTENTIAL.
FT TRANSMEM 605 625 POTENTIAL.
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 491 491 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 638 638 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 643 AA; 70569 MW; A5699FB6C556FF58 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 643;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 28 LIGAVNL 34
| | | | |
Db 366 LIGAVNL 372

Query Match 2.3%; Score 7; DB 1; Length 688;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 28 LIGAVNL 34
| | | | |
Db 366 LIGAVNL 372

PYS2_PSEAE
ID PYS2_PSEAE STANDARD; PRT; 688 AA.
AC Q06584;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pycocin S2 (EC 3.1.1.-) (Killer protein).
GN PYS2 OR PA1150.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-17.
RC STRAIN=PAO;
RX MEDLINE=93259934; PubMed=8491711;
RA Sano Y., Matsui H., Kobayashi M., Kageyama M.;
RT "Molecular structures and functions of pycocins S1 and S2 in
Pseudomonas aeruginosa.";
RL J. Bacteriol. 175:2907-2916(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong K.K.-S., Wu Z., Paulsen I.T.,
Reizer J., Saler M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- FUNCTION: CAUSES BREAKDOWN OF CHROMOSOMAL DNA AS WELL AS COMPLETE
INHIBITION OF LIPID SYNTHESIS IN SENSITIVE CELLS.
CC -!- SUBUNIT: PURIFIED PYCOCIN S2 MAKES UP A COMPLEX OF THE TWO (LARGE
AND SMALL) PROTEINS. THE LARGE PROTEIN, BUT NOT THE PYCOCIN
COMPLEX, SHOWS IN VITRO DNASE ACTIVITY.
CC -!- MISCELLANEOUS: PYCOCINS CONTAIN N-TERMINAL RECEPTOR-BINDING DOMAIN,
TRANSLLOCATION DOMAIN AND C-TERMINAL DNASE DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE NUCLEASE FAMILY OF COLICINS AND
PYOSINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; D12708; BAA02203.1; --
DR EMBL; AE004545; AAG04539.1; --
DR PIR; C36907; C36907.
DR PIR; D83501; D83501.
DR HSP; Q47112; 7CEI.
DR InterPro; IPR002711; HNH.
DR InterPro; IPR003615; HNH_nuc.
DR InterPro; IPR003060; Pycocin_killer.
DR Pfam; PF01844; HNH; 1.
DR PRINTS; PR01300; PYCOKILLER.
DR SMART; SM00507; HNHc; 1.
KW Antibiotic; Bacteriocin; Hydrolase; Endonuclease; Zinc; Metal-binding;
Complete proteome.
FT INIT MET 0
FT METAL 655 655 ZINC (BY SIMILARITY).
FT METAL 680 680 ZINC (BY SIMILARITY).
FT METAL 684 684 ZINC (BY SIMILARITY).
FT CONFLICT 604 604 R -> RR (IN REF.1)
SQ SEQUENCE 688 AA; 73722 MW; 1EB45076A75352F1 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 688;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 28 LIGAVNL 34
| | | | |
Db 366 LIGAVNL 372

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 146 VKAVPV 152
|||||
Db 451 VKAVPV 457

RESULT 32
NCPR ASPNG
ID NCPR ASPNG STANDARD; PRT; 693 AA.
AC Q00141;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE NADPH-cytochrome P450 reductase (EC 1.6.2.4) (CPR) (P450R).
GN CPRA.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A., AND ENZYME ACTIVITY.
RC STRAIN=ATCC 9029 / N204;
RX MEDLINE=95374684; PubMed=7646819;
RA van den Brink H.J.M., van Zeijl C.M.J., Brons J.F.,
RT "Cloning and characterization of the NADPH cytochrome P450
oxoreductase gene from the filamentous fungus Aspergillus niger";
RL DNA Cell Biol. 14:719-729(1995).
RN [2]
RP CHARACTERIZATION.
RC STRAIN=T18.5;
RX MEDLINE=21478854; PubMed=11594739;
RA Faber B.W., van Gorcom R.F.M., Duine J.A.;
RT "Purification and characterization of benzoate-para-hydroxylase, a
cytochrome P450 (CYP53A1), from Aspergillus niger";
RL Arch. Biochem. Biophys. 394:245-254(2001).
CC -!- FUNCTION: This enzyme is required for electron transfer from NADP
to cytochrome P450 in microsomes. It can also provide electron
transfer to heme oxygenase and cytochrome B5.
CC -!- CATALYTIC ACTIVITY: NADPH + 2 ferricytochrome = NADP(+) + 2
ferricytochrome.
CC -!- COFACTOR: FAD and FMN.
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum. Anchored to the ER
membrane by its N-terminal hydrophobic region (By similarity).
CC -!- INDUCTION: By benzoic acid (BA).
CC -!- SIMILARITY: THE FAD DOMAIN APPEARS HOMOLOGOUS WITH FERREDOXIN
NADP+ REDUCTASE, NADH CYTOCHROME B5 REDUCTASE, AND GLUTATHIONE
REDUCTASE.
CC -!- SIMILARITY: Contains 1 flavodoxin-like domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; Z26938; CAAB1550.1; -
DR PIR; S38427; S38427.
DR HSSP; P16435; 1B1C.
DR GO; GO:0005792; C:mitosome; IDA.
DR GO; GO:0008246; F:electron transfer flavoprotein; IDA.
DR GO; GO:0016491; F:oxidoreductase activity; IDA.
DR GO; GO:0006118; P:electron transport; IDA.
DR InterPro; IPR003097; FAD binding.
DR InterPro; IPR001094; Flavodoxin-like.
DR InterPro; IPR001226; Flavodoxin.
DR InterPro; IPR001709; FPN_cyt_reductase.
DR InterPro; IPR001433; Oxred_FAD/NAD(P).
DR Pfam; PF00667; FAD_binding_1; 1.
DR Pfam; PF00258; flavodoxin; 1.

DR Pfam; PF00175; NAD binding_1; 1.
DR PRINTS; PR00369; FLAVODOXIN.
DR PRINTS; PR00371; FPNCR.
DR PROSITE; PS50902; FLAVODOXIN-LIKE; 1.
KW Oxidoreductase; Flavoprotein; FMN; FAD; NADP; Endoplasmic reticulum;
KW Transmembrane.
FT TRANSMEM 9 31
FT DOMAIN 66 220
FT NP_BIND 165 196
FT NP_BIND 316 327
FT NP_BIND 445 456
FT NP_BIND 545 563
FT NP_BIND 640 656
SQ SEQUENCE 693 AA; 77072 MW; 1BEF35D4FC767674 CRC64;
Query Match 2.3%; Score 7; DB 1; Length 693;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 250 LVLAVAL 256
|||||
Db 9 LVLAVAL 15

RESULT 33
TCM2 CHICK
ID TCM2 CHICK STANDARD; PRT; 697 AA.
AC Q01841;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protein-glutamine gamma-glutamyltransferase (EC 2.3.2.13) (Tissue
transglutaminase) (TGase C) (TGC) (TG(C)) (Transglutaminase 2).
GN TG12.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Erythrocyte;
RX MEDLINE=93028551; PubMed=1357669;
RA Weraachakul-Boonmark N., Jeong J.M., Murthy S.N.P., Engel J.D.,
RT Lorand L.;
RT "Cloning and expression of chicken erythrocyte transglutaminase";
RL Proc. Natl. Acad. Sci. U.S.A. 89:9804-9808(1992).
CC -!- FUNCTION: CATALYZES THE CROSS-LINKING OF PROTEINS AND THE
CONJUGATION OF POLYAMINES TO PROTEINS.
CC -!- CATALYTIC ACTIVITY: Protein glutamine + alkylamine = protein N(5)-
alkylglutamine + NH(3).
CC -!- COFACTOR: Binds 1 calcium ion per subunit (By similarity).
CC -!- SUBUNIT: Monomer.
CC -!- TISSUE SPECIFICITY: PREDOMINATES IN MATURE ERYTHROCYTES. ALSO
FOUND IN KIDNEY AND CARDIAC MUSCLE.
CC -!- SIMILARITY: BELONGS TO THE TRANSGLUTAMINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; L02270; AAA49104.1; -
DR PIR; A47203; A47203.
DR HSSP; P00488; 1GGU.
DR InterPro; IPR001102; Gluttransfg.
DR InterPro; IPR002931; Trnsglutase_like.
DR Pfam; PF01841; transglut_core; 1.
DR Pfam; PF00927; transglutamin_C; 2.
DR Pfam; PF00868; transglutamin_N; 1.

```
DR SMART; SM00460; TGC; 1.
DR PROSITE; PS00547; TRANSGLUTAMINASES; 1.
KW Transferrase; Acyltransferase; Calcium-binding.
FT INIT MET 0 0
FT MOD RES 1 1 BLOCKED.
FT ACT_SITE 286 286 BY SIMILARITY.
FT ACT_SITE 344 344 BY SIMILARITY.
FT ACT_SITE 367 367 BY SIMILARITY.
FT METAL 407 407 CALCIUM (BY SIMILARITY).
FT METAL 409 409 CALCIUM (BY SIMILARITY).
FT METAL 455 455 CALCIUM (BY SIMILARITY).
FT METAL 460 460 CALCIUM (BY SIMILARITY).
SQ SEQUENCE 697 AA; 78608 MW; FE9916CB8442643 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 697;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 RLRLCAR 14
DB 514 RLRLCAR 520

RESULT 34
YJYIY_ECOLI
ID YJYIY_ECOLI STANDARD; PRT; 721 AA.
AC P39396;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yjiY.
GN YJYIY OR B4354.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -!- SIMILARITY: BELONGS TO THE CSTA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14003; AAA97251.1; -.
DR EMBL; AE000506; AAC77310.1; -.
DR PIR; S56580; S56580.
DR EcoGene; EG12586; YjiY.
DR InterPro; IPR003706; Csta.
DR Pfam; PF02554; Csta; 1.
KW Hypothetical protein; Transmembrane; Inner membrane;
FT TRANSMEM 11 31 POTENTIAL.
FT TRANSMEM 36 56 POTENTIAL.
FT TRANSMEM 94 114 POTENTIAL.
FT TRANSMEM 125 145 POTENTIAL.
FT TRANSMEM 169 189 POTENTIAL.
FT TRANSMEM 197 217 POTENTIAL.
FT TRANSMEM 228 248 POTENTIAL.
FT TRANSMEM 263 283 POTENTIAL.
```

```
FT TRANSMEM 288 308 POTENTIAL.
FT TRANSMEM 332 352 POTENTIAL.
FT TRANSMEM 380 400 POTENTIAL.
FT TRANSMEM 469 489 POTENTIAL.
FT TRANSMEM 491 511 POTENTIAL.
FT TRANSMEM 537 557 POTENTIAL.
FT TRANSMEM 574 594 POTENTIAL.
FT TRANSMEM 602 622 POTENTIAL.
FT TRANSMEM 674 694 POTENTIAL.
SQ SEQUENCE 721 AA; 77857 MW; D341DB9C1988C2C8 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 721;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 LAVIALI 259
DB 182 LAVIALI 188

RESULT 35
SC18_YEAST
ID SC18_YEAST STANDARD; PRT; 758 AA.
AC P18759;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Vesicular-fusion protein SEC18.
GN SEC18 OR YBR080C OR YBR0736.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89039834; PubMed=3054509;
RA Eakle K.A., Bernstein M., Emr S.D.;
RT "Characterization of a component of the yeast secretion machinery:
RT identification of the SEC18 gene product.";
RL Mol. Cell. Biol. 8:4098-4109(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C;
RX MEDLINE=95076715; PubMed=7985423;
RA van der Aart Q.J.M., Barthe C., Doignon F., Aigle M., Crouzet M.,
RA Steensma H.Y.;
RT "Sequence analysis of a 31 kb DNA fragment from the right arm of
RT Saccharomyces cerevisiae chromosome II.";
RL Yeast 10:959-964(1994).
RN [3]
RP SEQUENCE OF 566-758 FROM N.A.
RX STRAIN=W303;
RA Vornlocher H.-P., Hanachi P., Hershey J.W.B.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR VESICLE-MEDIATED TRANSPORT. CATALYZES THE
CC FUSION OF TRANSPORT VESICLES WITHIN THE GOLGI CISTERNAE. IS ALSO
CC REQUIRED FOR TRANSPORT FROM THE ENDOPLASMIC RETICULUM TO THE GOLGI
CC STACK. SEEM TO FUNCTION AS A FUSION PROTEIN REQUIRED FOR THE GOLGI
CC DELIVERY OF CARGO PROTEINS TO ALL COMPARTMENTS OF THE GOLGI STACK
CC INDEPENDENT OF VESICLE ORIGIN.
CC -!- SUBUNIT: BINDS TO SEC17.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M20662; AAA35030.1; -.
```

```

DR EMBL; X76294; CAA53939.1; -.
DR EMBL; Z35949; CAA85025.1; -.
DR EMBL; AF004912; AAB82417.1; -.
DR PIR; S45477; S45477.
DR PDB; 1CR5; 22-DEC-99.
DR SGD; S0000284; SEC18.
DR GO; GO:0004002; C:extrinsic to plasma membrane; IDA.
DR GO; GO:0004002; P:adenosinetriphosphatase activity; IDA.
DR GO; GO:0000011; P:vacuole inheritance; IDA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase.
DR InterPro; IPR003960; AAA sub.
DR InterPro; IPR003338; ATPaseVAT N.
DR InterPro; IPR004201; Cdc48_2.
DR Pfam; PF00004; AAA; 2.
DR Pfam; PF02933; Cdc48_2; 1.
DR Pfam; PF02359; Cdc48_N; 1.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00674; AAA; 1.
KW Transport; Protein transport; Endoplasmic reticulum; Golgi stack;
KW ATP-binding; Repeat; 3D-structure.
FT NP_BIND 281 288 ATP (POTENTIAL).
FT NP_BIND 564 571 ATP (POTENTIAL).
FT CONFLICT 381 381 MISSING (IN REF. 1).
SQ SEQUENCE 758 AA; 84056 MW; AD6CE77EA3674B3C CRC64;

Query Match 2.3%; Score 7; DB 1; Length 758;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 32 VNLKSSN 38
Db 214 VNLKSSN 220

RESULT 36
CC16 YEAST
ID CC16 YEAST STANDARD; PRT; 840 AA.
AC P09798;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Cell division control protein 16.
GN CDC16 OR YKL022C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R.B.Wickner 1385;
RX MEDLINE=88040465; PubMed=2823230;
RA Icho T., Wickner R.B.;
RT "Metal-binding, nucleic acid-binding finger sequences in the CDC16
gene of Saccharomyces cerevisiae."
RL Nucleic Acids Res. 15:8439-8450(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA Rieger M.;
RN Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP TPR REPEATS.
RX MEDLINE=90124639; PubMed=2404612;
RA Sikorski R.S., Boguski M.S., Goebel M., Hieter P.A.;
RT "A repeating amino acid motif in CDC23 defines a family of proteins
and a new relationship among genes required for mitosis and RNA
synthesis."
RL Cell 60:307-317(1990).
RN [4]
RP SUBUNITS.
RX MEDLINE=95009933; PubMed=7925276;
RA Lamb J.R., Michaud W.A., Sikorski R.S., Hieter P.A.;
RT "Cdc16p, Cdc23p and Cdc27p form a complex essential for mitosis."

```

```

RL EMBL; J. 13:4321-4328(1994).
CC -!- FUNCTION: EXACT FUNCTIONS NOT KNOWN. REQUIRED FOR CHROMOSOME
CC SEGREGATION. MUTATIONS IN CDC16 CAUSE CELLS TO ARREST UNIFORMLY
CC AT G2/M AFTER DNA REPLICATION BUT PRIOR TO MITOTIC SPINDLE
CC ELONGATION.
CC -!- SUBUNIT: CDC16, CDC23 AND CDC27 FORMS A MACROMOLECULAR COMPLEX.
CC -!- SIMILARITY: Contains 10 TPR repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X06165; CAA29521.1; -.
DR EMBL; Z28022; CAA81857.1; -.
DR PIR; A27832; A27832.
DR SGD; S0001505; CDC16.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 7.
DR SMART; SM00028; TPR; 6.
KW Cell division; Cell cycle; Mitosis; Repeat; TPR repeat;
KW Nuclear protein.
FT REPEAT 263 295 TPR 1.
FT REPEAT 296 329 TPR 2.
FT REPEAT 359 392 TPR 3.
FT REPEAT 497 530 TPR 4.
FT REPEAT 531 564 TPR 5.
FT REPEAT 565 598 TPR 6.
FT REPEAT 599 632 TPR 7.
FT REPEAT 633 666 TPR 8.
FT REPEAT 674 707 TPR 9.
FT REPEAT 708 741 TPR 10.
SQ SEQUENCE 840 AA; 94991 MW; A096B34441083488 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 840;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 33 NLKSSNR 39
Db 30 NLKSSNR 36

RESULT 37
RPS2 ARATH
ID RPS2 ARATH STANDARD; PRT; 909 AA.
AC Q42484; O82096; Q8L3R0; Q8L3W3; Q8L4X9; Q8L4Y0; Q8L5B7; Q8L5B3;
AC Q8LKZ8; Q8LKZ9; Q8LL00; Q8LL01; Q9ASP5;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Disease resistance protein RPS2 (Resistance to Pseudomonas syringae
protein 2).
GN RPS2 OR Atg26090 OR F20B18.200.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsie.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=94377978; PubMed=8091210;
RA Bent A.F., Kunkel B.N., Dahlbeck D., Brown K.L., Schmidt R.,
RA Giraudet J., Leung J., Staskiewicz B.J.;
RT "RPS2 of Arabidopsis thaliana: a leucine-rich repeat class of plant
disease resistance genes."
RL Science 265:1856-1860(1994).
RN [2]

```

RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=95007758; PubMed=7923358;
RA Mindrinos M., Katagiri F., Yu G.-L., Ausubel F.M.;
RT "The A. thaliana disease resistance gene RPS2 encodes a protein
RL containing a nucleotide-binding site and leucine-rich repeats";
RN Cell 78:1089-1099(1994).
[3]
RP SEQUENCE FROM N.A., AND VARIANTS.
RC STRAIN=cv. PO-1;
RX STRAIN=cv. PO-1; PubMed=11333251;
RA Banerjee D., Zhang X., Bent A.F.;
RT "The leucine-rich repeat domain can determine effective interaction
RL between RPS2 and other host factors in Arabidopsis RPS2-mediated
RN disease resistance";
RN Genetics 158:439-450(2001).
[4]
RP SEQUENCE FROM N.A., AND VARIANTS.
RC STRAIN=cv. Ab-7, cv. Ang-0, cv. Bg-4, cv. Bla-2, cv. Bur-0, cv. C2-1,
RX cv. CO-1, cv. Ct-1, cv. CVI-0, cv. D2-9, cv. Fm-17, cv. G2-1,
RC cv. Gott-20, cv. Gr-6, cv. Hs-12, cv. Kas-1, cv. KNO2, cv. Mt-0,
RC cv. Tamm-17, cv. Tsu-0, cv. PO-1, cv. Pog-0, cv. Pu-8, cv. RLD,
RC cv. YO-0, cv. Wu-0, and cv. Zu-0;
RX MEDLINE=22505406; PubMed=12618410;
RA Mauricio R., Stahl E.A., Korves T., Tian D., Kreitman M.,
RA Bergelson J.;
RT "Natural selection for polymorphism in the disease resistance gene
RL rps2 of Arabidopsis thaliana";
RN Genetics 163:735-746(2003).
[5]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambert R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Anorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Delaney M., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weitzens I., Voet M., Baatiens I., Aert R., Defoor E.,
RA Weitzenecker T., Bothe G., Ransperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Mooljman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buyssehaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIlroy K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloecker H., Scharfe M., Grimm M., Lohner T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Farman B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chedori F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threiden J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Krawar J., Fulton J., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antonolu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,

RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RL thaliana";
RN Nature 402:769-777(1999).
[6]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinozaki K.;
RT "Arabidopsis thaliana full-length cDNA";
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
[7]
RP SEQUENCE OF 184-352 FROM N.A.
RC STRAIN=cv. Nd-1;
RX MEDLINE=98335218; PubMed=9670562;
RA Spelman E., Bouchez D., Holub E.B., Beynon J.L.;
RT "Disease resistance gene homologs correlate with disease resistance
RL loci of Arabidopsis thaliana";
RN Plant J. 14:467-474(1998).
[8]
RP SUBCELLULAR LOCATION, AND MUTAGENESIS OF ILE-353.
RX MEDLINE=97140359; PubMed=8986840;
RA Leister R.T., Ausubel F.M., Katagiri F.;
RT "Molecular recognition of pathogen attack occurs inside of plant cells
RL in plant disease resistance specified by the Arabidopsis genes RPS2
RN and RPM1";
RN Proc. Natl. Acad. Sci. U.S.A. 93:15497-15502(1996).
[9]
RP IDENTIFICATION IN A COMPLEX CONTAINING AVRPT2 AND AVR8.
RX MEDLINE=20307433; PubMed=10849351;
RA Leister R.T., Katagiri F.;
RT "A resistance gene product of the nucleotide binding site -- leucine
RL rich repeats class can form a complex with bacterial avirulence
RN proteins in vivo";
RN Plant J. 22:345-354(2000).
[10]
RP MUTAGENESIS OF 38-LEU-THR-40, LYS-188, THR-189, THR-190 AND
RX 262-GLU-GLU-263.
RA MEDLINE=21066938; PubMed=11148296;
TAO Y., Yuan F., Leister R.T., Ausubel F.M., Katagiri F.;
RT "Mutational analysis of the Arabidopsis nucleotide binding site-
RL leucine-rich repeat resistance gene RPS2";
RN Plant Cell 12:2541-2554(2000).
[11]
RP MUTANTS 204C, 205C, 206C, 209C, 210C AND 211C.
RX MEDLINE=21071235; PubMed=11204781;
RA Axtell M.J., McNeill T.W., Mudgett M.B., Hsu C.S., Staskiewicz B.J.;
RT "Mutational analysis of the Arabidopsis RPS2 disease resistance gene
RL and the corresponding pseudomonas syringae avrP2 avirulence gene";
RN Mol. Plant Microbe Interact. 14:181-188(2001).
[12]
RP SUBCELLULAR LOCATION, AND INTERACTION WITH AVRPT2 AND RIN4.
RX MEDLINE=22469031; PubMed=12581526;
RA Axtell M.J., Staskiewicz B.J.;
RT "Initiation of RPS2-specified disease resistance in Arabidopsis is
RL coupled to the AvrP2-directed elimination of RIN4";
RN Cell 112:369-377(2003).
[13]
RP FUNCTION.
RX MEDLINE=22469032; PubMed=12581527;
RA Mackey D., Belkadir Y., Alonso J.M., Ecker J.R., Dangl J.L.;
RT "Arabidopsis RIN4 is a target of the type III virulence effector
RL AvrP2 and modulates RPS2-mediated resistance";
RN Cell 112:379-389(2003).
CC -1- FUNCTION: Disease resistance (R) protein that specifically
CC recognizes the AvrP2 type III effector avirulence protein from
CC pseudomonas syringae. Resistance proteins guard the plant against
CC pathogens that contain an appropriate avirulence protein via an
CC indirect interaction with this avirulence protein. That triggers a
CC defense system including the hypersensitive response, which

restricts the pathogen growth. Acts via its interaction with RIN4, and probably triggers the plant resistance when RIN4 is degraded by AvrRpt2.

-!- SUBUNIT: Interacts indirectly with RIN4. Found in a complex with AVRrpt2 and AvrB.

-!- SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated.

-!- DOMAIN: The LRR repeats probably act as specificity determinant of pathogen recognition (By similarity).

-!- DOMAIN: The Leucine-zipper domain is essential for the resistance to AvrRpt2; the cultivars that do not display resistance showing specific variations in this region.

-!- POLYMORPHISM: The polymorphism between the different cultivars influence the specificity to the pathogen recognition. In cv. Po.1, KNO2, Bg-4 and Zu-0, RPS2 does not confer resistance to AVRrpt2.

-!- SIMILARITY: Belongs to the disease resistance NB-LRR family.

-!- SIMILARITY: Contains 6 leucine-rich (LRR) repeats.

-!- SIMILARITY: Contains 1 NB-ARC domain.

-!- DATABASE: NAME=NB-LRRs;

NOTE=Functional and comparative genomics of disease resistance gene homologs;

WWW="http://niblr.ucdavis.edu".

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

Query Match 2.3%; Score 7; DB 1; Length 909;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 LALITLG 262
|||||
350 LALITLG 356

Db

RESULT 38

SCA4 RICFE

ID SCA4 RICFE STANDARD; PRT; 981 AA.

AC Q9AJ37;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen) (Protein PS 120) (Fragment).

DE (Protein PS 120) (Fragment).

GN SCA4 OR D. (Rickettsia azadi).

OS Rickettsia felis (Rickettsia azadi).

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;

OC Rickettsiaceae; Rickettsieae; Rickettsia.

OX NCBI_TaxID=42862;

RN [1]

RP SEQUENCE FROM N.A.

RA Sekeyova Z., Roux V., Raoult D.;

RT "Phylogenetic analysis of Rickettsia spp. by comparing the sequence of gene D coding for an intracytoplasmic protein.";

RL Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF196973; AAK31304.1; -.

KW Antigen.

FT NON_TER 1 1

FT NON TER 981 981
SQ SEQUENCE 981 AA; 107514 MW; 7F18F421E2C262E1 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 981;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 DSQTSDDP 63
|||||
57 DSQTSDDP 63

Db

RESULT 39

K6P1 CANAL

ID K6P1 CANAL STANDARD; PRT; 987 AA.

AC O94201;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE 6-phosphofructokinase alpha subunit (EC 2.7.1.11) (Phosphofructokinase 1) (Phosphohexokinase) (6PF-1-K alpha subunit) (CaPFK1).

GN PFK1.

OS Candida albicans (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; mitosporic Saccharomycetales; Candida.

OX NCBI_TaxID=5476;

RN [1]

RP SEQUENCE FROM N.A., AND CHARACTERIZATION.

RC STRAIN=ATCC 10231;

RC MEDLINE=99191702; PubMed=10091602;

RA Loeberig A., Kirchhuth L., Ernst J.F., Heinisch J.J.;

RT "Genetic and biochemical characterization of phosphofructokinase from the opportunistic pathogenic yeast Candida albicans.";

RL Eur. J. Biochem. 260:217-226(1999).

CC -!- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-fructose 1,6-bisphosphate.

CC -!- ENZYME REGULATION: Allosterically inhibited by ATP and activated by AMP and fructose 2,6-bisphosphate.

CC -!- PATHWAY: Key control step of glycolysis.

CC -!- SUBUNIT: Heteroctamer of 4 alpha and 4 beta chains (By similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -!- SIMILARITY: Belongs to the phosphofructokinase family. Two domains subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AJ007638; CAB38868.1; -.

DR HSSP; P00512; 3PFK.

DR InterPro; IPR000023; Ppfuckinase.

DR Pfam; PF00365; PFK; 2.

DR PRINTS; PR00476; PPFCKTKINASE.

DR PRODOM; PD000707; Ppfuckinase; 2.

DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 1.

KW Kinase; Transferase; Glycolysis; Repeat; Allosteric enzyme.

SQ SEQUENCE 987 AA; 108597 MW; 4EE5BAB6D02349FA CRC64;

Query Match 2.3%; Score 7; DB 1; Length 987;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 QEMEYVD 239
|||||
91 QEMEYVD 97

Db

RESULT 40


```
SCA4_RICSL
ID SCA4_RICSL STANDARD; PRT; 991 AA.
AC Q9AJ77;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (Protein PS 120) (Fragment).
GN SCA4 OR D.
OS Rickettsia sibirica.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=35793;
RN [1]
RP SEQUENCE FROM N.A.
RA Sekeyova Z., Roux V., Raoult D.;
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
'gene D' coding for an intracytoplasmic protein.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF155057; AAK30688.1; -
KW Antigen.
FT NON TER 1
FT NON TER 991
SQ SEQUENCE 991 AA; 108564 MW; 3F499954933D715C CRC64;

Query Match 2.3%; Score 7; DB 1; Length 991;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 DSQTSDP 63
Db 67 DSQTSDP 73

RESULT 41
SCA4_RICAF
ID SCA4_RICAF STANDARD; PRT; 1011 AA.
AC Q9AJ83;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (Protein PS 120) (Fragment).
GN SCA4 OR D.
OS Rickettsia africana.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=35788;
RN [1]
RP SEQUENCE FROM N.A.
RA Sekeyova Z., Roux V., Raoult D.;
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
'gene D' coding for an intracytoplasmic protein.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF151725; AAK30683.1; -
KW Antigen.
FT NON TER 1
FT NON TER 1011
SQ SEQUENCE 1011 AA; 110607 MW; 0169A06981BD5D08 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 1011;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 DSQTSDP 63
Db 67 DSQTSDP 73

RESULT 42
SCA4_RICWN
ID SCA4_RICWN STANDARD; PRT; 1011 AA.
AC Q9AJ82;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (Protein PS 120) (Fragment).
GN SCA4 OR D.
OS Rickettsia mongolotimonae.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=45261;
RN [1]
RP SEQUENCE FROM N.A.
RA Sekeyova Z., Roux V., Raoult D.;
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
'gene D' coding for an intracytoplasmic protein.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF151724; AAK30682.1; -
KW Antigen.
FT NON TER 1
FT NON TER 1011
SQ SEQUENCE 1011 AA; 110491 MW; 60F7C4DD58D438D CRC64;

Query Match 2.3%; Score 7; DB 1; Length 1011;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 DSQTSDP 63
Db 67 DSQTSDP 73

RESULT 43
SCA4_RICSL
ID SCA4_RICSL STANDARD; PRT; 1012 AA.
AC Q9AJ80;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (Protein PS 120) (Fragment).
GN SCA4 OR D.
OS Rickettsia slovaca.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
```

```
OC Rickettsiaceae; Rickettsiales; Rickettsia.
OX NCBI_TaxID=35794;
RN [1]
RP SEQUENCE FROM N.A.
RA Sekeyova Z., Roux V., Raoult D.;
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
RL 'gene D' coding for an intracytoplasmic protein.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF155054; AAK30685.1; -
KW Antigen.
FT NON_TER 1
FT NON_TER 1012
FT NON_TER 1013
SQ SEQUENCE 1012 AA; 110694 MW; E55A4D9077DDB2D1 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 1012;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 DSQTSDDP 63
DB 67 DSQTSDDP 73

RESULT 44.
SCA4 RICRH STANDARD; PRT; 1013 AA.
AC O9AJ81;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (Protein PS 120) (Fragment).
GN SCA4 OR D.
OS Rickettsia rhipicephali.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiales; Rickettsia.
OX NCBI_TaxID=33992;
RN [1]
RP SEQUENCE FROM N.A.
RA Sekeyova Z., Roux V., Raoult D.;
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
RL 'gene D' coding for an intracytoplasmic protein.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF155053; AAK30684.1; -
KW Antigen.
FT NON_TER 1
FT NON_TER 1013
SQ SEQUENCE 1013 AA; 110550 MW; 856E98912315D102 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 1013;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 57 DSQTSDDP 63
DB 67 DSQTSDDP 73

RESULT 45
SCA4 RICJA STANDARD; PRT; 1018 AA.
AC O9AJ79; O06654;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (Protein PS 120) (rps120).
GN SCA4 OR D.
OS Rickettsia japonica.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiales; Rickettsia.
OX NCBI_TaxID=35790;
RN [1]
RP SEQUENCE FROM N.A.
RA Sekeyova Z., Roux V., Raoult D.;
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
RL 'gene D' coding for an intracytoplasmic protein.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB003696; BAA20142.1; -
DR EMBL; AF155055; AAK30686.1; -
DR PIR; T30853; T30853.
KW Antigen.
FT CONFLICT 8 8 N -> D (IN REF. 2).
FT CONFLICT 449 449 K -> E (IN REF. 2).
FT CONFLICT 693 693 S -> G (IN REF. 2).
SQ SEQUENCE 1018 AA; 111147 MW; F41F015392671BAA CRC64;

Query Match 2.3%; Score 7; DB 1; Length 1018;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 DSQTSDDP 63
DB 74 DSQTSDDP 80

RESULT 46
SCA4 RICCN STANDARD; PRT; 1022 AA.
AC Q52658;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
DE (Protein PS 120) (RCA).
GN SCA4 OR RC0667.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
```

CC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=941156485; PubMed=8112862;
RA Schuenke K.W., Walker D.H.;
RT "Cloning, sequencing, and expression of the gene coding for an
RL antigenic 120-kilodalton protein of Rickettsia conorii";
RN Infect. Immun. 62:904-909(1994).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
RL Science 293:2093-2098(2001).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U01133; AAA18636.1; -;
DR EMBL; AF008626; AAL03205.1; ALT INIT.
KW Antigen; Complete proteome.
SQ SEQUENCE 1022 AA; 111802 MW; 8F1042542B7B9982 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 1022;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 57 DSQTSDDP 63
Db 74 DSQTSDDP 80
|||||

RESULT 47

ID ITA3 MOUSE STANDARD; PRT; 1053 AA.
AC Q62470; Q08441; Q08442;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-SEP-2003 (Rel. 41, Last annotation update)
DE Integrin alpha-3 precursor (Galactoprotein B3) (GAPB3) (VLA-3 alpha
DE chain) (CD49c).
GN ITGA3.
OS Mus musculus. (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BAUB/c;
RX MEDLINE=95279462; PubMed=7759572;
RA Takeuchi K., Hirano K., Tuji T., Osawa T., Irimura T.;
RT "cDNA cloning of mouse VLA-3 alpha subunit";
RL J. Cell. Biochem. 57:371-377(1995).
RN [2]
RP SEQUENCE OF 913-1053 FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=92052235; PubMed=1946438;
RA Tamura R.N., Cooper H.M., Collo G., Quaranta V.;
RT "Cell type-specific integrin variants with alternative alpha chain
RT cytoplasmic domains";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10183-10187(1991).
CC -!- FUNCTION: INTEGRIN ALPHA-3/BETA-1 IS A RECEPTOR FOR FIBRONECTIN,
CC LAMININ, COLLAGEN, EPILIGRIN AND THROMBOSPONDIN.

CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
CC SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A
CC DISULFIDE BOND. ALPHA-3 ASSOCIATES WITH BETA-1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Events=Alternative splicing; Named isoforms=2;
CC Name=Alpha-3a;
CC IsoId=Q62470-1; Sequence=Displayed;
CC Name=Alpha-3b;
CC IsoId=Q62470-2; Sequence=VSP 002722;
CC IsoId=Q62470-1; Sequence=VSP 002722;
CC -!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D13867; BAA02980.1; -;
DR EMBL; S66292; AAB20356.2; -;
DR EMBL; S66294; AAB20357.2; -;
DR PIR; I55534; I55534.
DR HSSP; P11215; IABX.
DR MGD; MGI:96602; Itga3.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; EG-GAP; 3.
DR Pfam; PF00357; Integrin_A; 1.
DR PRINTS; PRO1185; INTEGRIN.
DR SMART; SMO0191; Int_alpha; 5.
DR PROSITE; PS00242; INTEGRIN ALPHA; 1.
KW Integrin; Cell adhesion; Glycoprotein; Transmembrane; Signal; Repeat;
KW Alternative splicing.
FT SIGNAL 1 32 BY SIMILARITY.
FT CHAIN 33 1053 INTEGRIN ALPHA-3.
FT CHAIN 33 874 INTEGRIN ALPHA-3 HEAVY CHAIN (POTENTIAL).
FT CHAIN 878 1053 INTEGRIN ALPHA-3 LIGHT CHAIN (POTENTIAL).
FT DOMAIN 33 993 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 994 1021 POTENTIAL.
FT DOMAIN 1022 1053 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 49 462 7 X APPROXIMATE REPEATS.
FT REPEAT 49 94 I.
FT REPEAT 120 165 II.
FT REPEAT 195 227 III.
FT REPEAT 246 280 IV.
FT REPEAT 305 346 V.
FT REPEAT 367 403 VI.
FT REPEAT 427 462 VII.
FT DISULFID 94 103 BY SIMILARITY.
FT DISULFID 140 162 BY SIMILARITY.
FT DISULFID 185 197 BY SIMILARITY.
FT DISULFID 486 491 BY SIMILARITY.
FT DISULFID 497 551 BY SIMILARITY.
FT DISULFID 616 622 BY SIMILARITY.
FT DISULFID 695 704 BY SIMILARITY.
FT DISULFID 848 906 INTERCHAIN (BY SIMILARITY).
FT DISULFID 913 918 BY SIMILARITY.
FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 512 512 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 574 574 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 606 606 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 657 657 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 699 699 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 843 843 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 859 859 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 925 925 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 928 928 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 937 937 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 971 971 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 1023 1053 IMPKHAVRIREDRYPPPGSTLPTKKHW (in isoform
FT PT

```
FT FT CONFLICT 975 975 W -> C (IN REF. 2).
FT FT CONFLICT 979 979 D -> N (IN REF. 2).
FT FT CONFLICT 1002 1002 G -> S (IN REF. 2).
FT FT CONFLICT 1019 1019 G -> D (IN REF. 2).
SQ SEQUENCE 1053 AA; 116745 MW; 6A58BFBDBA86D6E5 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 1053;
Best Local Similarity 100.0%; Pred. No. 1.2e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

QY 4 RRPRLR 10
Db 525 RRPRLR 531

RESULT 48
ID ITA3 CRISP STANDARD; PRT; 1066 AA.
AC P17852;
DT 01-NOV-1990 (Rel. 16, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Integrin alpha-3 precursor (Galactoprotein B3) (GAPB3) (VLA-3 alpha
DE chain) (CD49c).
GN ITA3.
OS Cricetidae sp. (Hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae.
OX NCBI_TaxID=36483;
RN [1] _TaxID=36483;
RP SEQUENCE FROM N.A. (ISOFORM ALPHA-3A), AND PARTIAL SEQUENCE.
RC TISSUE=Fibroblast;
RX MEDLINE=90216739; PubMed=1691184;
RA Tsuji T., Yamamoto F.-I., Miura Y., Takio K., Titani K., Pawar S.,
RA Osawa T., Hakomori S.-I.;
RT "Characterization through cDNA cloning of galactoprotein b3 (Gap b3),
RT a cell surface membrane glycoprotein showing enhanced expression on
RT oncogenic transformation. Identification of Gap b3 as a member of the
RT integrin superfamily."
RL J. Biol. Chem. 265:7016-7021(1990).
RN [2]

ALTERNATIVE SPLICING
RX MEDLINE=92052335; PubMed=1946438;
RA Tamura R.N., Cooper H.M., Collo G., Quaranta V.;
RT "Cell type-specific integrin variants with alternative alpha chain
RT cytoplasmic domains."
RL Proc. Natl. Acad. Sci. U.S.A. 88:10183-10187(1991).
CC -!- FUNCTION: INTEGRIN ALPHA-3/BETA-1 IS A RECEPTOR FOR FIBRONECTIN,
CC LAMININ, COLLAGEN, EPILIGRIN AND THROMBOSPONDIN.
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. THE ALPHA
CC SUBUNIT IS COMPOSED OF AN HEAVY AND A LIGHT CHAIN LINKED BY A
CC DISULFIDE BOND. ALPHA-3 ASSOCIATES WITH BETA-1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Alpha-3B;
CC IsoId=P17852-1; Sequence=Displayed;
CC Name=Alpha-3A;
CC IsoId=P17852-2; Sequence=VSP 002720;
CC Notes=No experimental confirmation available;
CC -!- PTM: ISOFORM ALPHA-3A, BUT NOT ISOFORM ALPHA-3B, IS PHOSPHORYLATED
CC ON SERINE RESIDUES (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
CC -!- SIMILARITY: Contains 7 FG-GAP repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC or send an email to license@isb-sib.ch).
CC EMBL; J05281; AAA56794.1; -.
DR HSSP; P11215; 1A8X
DR InterPro; IPR000413; Integrin_alpha
DR Pfam; PF01839; FG-GAP; 3.
DR PRINTS; PR01185; INTEGRINA.
DR SMART; SM00191; Int alpha; 5.
DR PROSITE; PS00242; INTEGRIN ALPHA; FALSE NEG.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
KW Signal; Phosphorylation; Repeat; Alternative splicing; Calcium.
FT SIGNAL 1 32
FT CHAIN 33 1066 INTEGRIN ALPHA-3.
FT CHAIN 33 872 INTEGRIN ALPHA-3 HEAVY CHAIN (POTENTIAL).
FT CHAIN 876 1066 INTEGRIN ALPHA-3 LIGHT CHAIN (POTENTIAL).
FT DOMAIN 33 991 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 992 1019 POTENTIAL.
FT DOMAIN 1020 1066 CYTOPLASMIC (POTENTIAL).
FT REPEAT 49 94 FG-GAP 1.
FT REPEAT 120 165 FG-GAP 2.
FT REPEAT 195 227 FG-GAP 3.
FT REPEAT 246 279 FG-GAP 4.
FT REPEAT 304 345 FG-GAP 5.
FT REPEAT 366 402 FG-GAP 6.
FT REPEAT 426 461 FG-GAP 7.
FT CA_BIND 315 323 POTENTIAL.
FT CA_BIND 378 386 POTENTIAL.
FT CA_BIND 439 447 POTENTIAL.
FT SITE 1017 1021 GFFKR MCTIF.
FT DISULFID 94 103 BY SIMILARITY.
FT DISULFID 140 162 BY SIMILARITY.
FT DISULFID 185 197 BY SIMILARITY.
FT DISULFID 485 490 BY SIMILARITY.
FT DISULFID 496 550 BY SIMILARITY.
FT DISULFID 615 621 BY SIMILARITY.
FT DISULFID 694 702 BY SIMILARITY.
FT DISULFID 846 904 INTERCHAIN (BY SIMILARITY).
FT DISULFID 911 916 BY SIMILARITY.
FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 500 500 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 511 511 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 605 605 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 656 656 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 697 697 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 841 841 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 923 923 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 935 935 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 969 969 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 1021 1066 PTRYVIRMPKVHVRIRREERYPPPGSTLPTKKHWVTSMQI
RDYV -> RARTALYEAKQKAEKMSQPSETERLTDDY
(in isoform Alpha-3A).
SQ SEQUENCE 1066 AA; 118550 MW; FD089C2431F57684 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 1066;
Best Local Similarity 100.0%; Pred. No. 1.2e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

QY 4 RRPRLR 10
Db 524 RRPRLR 530

RESULT 49
ID ITA3 HUMAN STANDARD; PRT; 1066 AA.
AC P26006;
DT 01-MAY-1992 (Rel. 22, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Integrin alpha-3 precursor (Galactoprotein B3) (GAPB3) (VLA-3 alpha
```

chain) (CD49c).
ITGA3.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORMS ALPHA-3A AND ALPHA-3B).
MEDLINE=92011866; PubMed=1655803;
Takada Y., Murphy E., Pil P., Chen C., Ginsberg M.H., Hemler M.E.;
"Molecular cloning and expression of the cDNA for alpha 3 subunit of
human alpha 3 beta 1 (VLA-3), an integrin receptor for fibronectin,
laminin, and collagen.";
J. Cell Biol. 115:257-266 (1991).
[2]
PARTIAL SEQUENCE FROM N.A. (ISOFORM ALPHA-3A).
TISSUE=Fibroblast;
MEDLINE=91331981; PubMed=1714443;
Tsuji T., Hakomori S.-I., Osawa T.;
"Identification of human galactoprotein b3, an oncogenic
transformation-induced membrane glycoprotein, as VLA-3 alpha
subunit: the primary structure of human integrin alpha 3.";
J. Biochem. 109:659-665 (1991).
[3]
SEQUENCE OF 33-46.
MEDLINE=87204112; PubMed=3033641;
Takada Y., Strominger J.L., Hemler M.E.;
"The very late antigen family of heterodimers is part of a
superfamily of molecules involved in adhesion and embryogenesis.";
Proc. Natl. Acad. Sci. U.S.A. 84:3239-3243 (1987).
[4]
ALTERNATIVE SPLICING, PHOSPHORYLATION, AND TISSUE SPECIFICITY.
MEDLINE=97265610; PubMed=9111516;
de Melker A.A., Sterk L.M., Delwel G.O., Fies D.L., Daams H.,
Weening J.J., Sonnenberg A.;
"The A and B variants of the alpha 3 integrin subunit: tissue
distribution and functional characterization.";
Lab. Invest. 76:547-563 (1997).
-!- FUNCTION: INTEGRIN ALPHA-3/BETA-1 IS A RECEPTOR FOR FIBRONECTIN,
LAMININ, COLLAGEN, EPLIGRIN AND THROMBOSPONDIN.
-!- SUBUNIT: Heterodimer of an alpha and a beta subunit. The alpha
subunit is composed of an heavy and a light chain linked by a
disulfide bond. Alpha-3 associates with beta-1. Interacts with
HP55.
-!- SUBCELLULAR LOCATION: Type I membrane protein.
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=Alpha-3B;
IsoId=P26006-1; Sequence=Displayed;
Name=Alpha-3A;
IsoId=P26006-2; Sequence=VSP 002721;
-!- TISSUE SPECIFICITY: ISOFORM ALPHA-3A IS WIDELY EXPRESSED. ISOFORM
ALPHA-3B IS EXPRESSED IN BRAIN AND HEART. IN BRAIN, BOTH ISOFORMS
ARE EXCLUSIVELY EXPRESSED ON VASCULAR SMOOTH MUSCLE CELLS, WHEREAS
IN HEART ISOFORM ALPHA-3A IS STRONGLY EXPRESSED ON VASCULAR SMOOTH
MUSCLE CELLS. ISOFORM ALPHA-3B IS DETECTED ONLY ON ENDOTHELIAL
VEIN CELLS.
-!- PTM: ISOFORM ALPHA-3A, BUT NOT ISOFORM ALPHA-3B, IS PHOSPHORYLATED
ON SERINE RESIDUES. PHOSPHORYLATION INCREASES AFTER PHORBOL 12-
MYRISTATE 13-ACETATE STIMULATION.
-!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
-!- SIMILARITY: Contains 7 FG-GAP repeats.
-!- DATABASE: NAME=PROW; NOTE=CD guide CD49c entry;
WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd49c.htm".

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; M59911; AAA36120.1; -;
DR EMBL; D01038; BAA00845.1; -;
DR PIR; A40021; A40021.
DR HSSP; P11215; 1A8X.
DR MIW; G05025; -;
DR GO; G0008305; C: Integrin complex; TAS.
DR GO; G0004895; F: cell adhesion receptor activity; TAS.
DR GO; G0007160; P: cell-matrix adhesion; TAS.
DR InterPro; IPR000413; Integrin_alpha.
DR Pfam; PF01839; FG-GAP 3.
DR PRINTS; PRO1185; INTEGRIN.
DR SMART; SM00191; Int_alpha; 5.
DR PROSITE; PS00242; INTEGRIN ALPHA; 1.
KW Integrin; Cell adhesion; Receptor; Glycoprotein; Transmembrane;
Signal; Phosphorylation; Repeat; Alternative splicing; Calcium.
FT SIGNAL 1 32
FT CHAIN 33 1066 INTEGRIN ALPHA-3.
FT CHAIN 33 872 INTEGRIN ALPHA-3 HEAVY CHAIN (POTENTIAL).
FT CHAIN 876 1066 INTEGRIN ALPHA-3 LIGHT CHAIN (POTENTIAL).
FT DOMAIN 33 991 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 992 1014 POTENTIAL.
FT DOMAIN 1015 1066 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1015 1021 INTERACTION WITH HP55.
FT REPEAT 49 94 FG-GAP 1.
FT REPEAT 120 165 FG-GAP 2.
FT REPEAT 195 227 FG-GAP 3.
FT REPEAT 246 279 FG-GAP 4.
FT REPEAT 304 345 FG-GAP 5.
FT REPEAT 366 402 FG-GAP 6.
FT REPEAT 426 461 FG-GAP 7.
FT CA_BIND 315 323 POTENTIAL.
FT CA_BIND 378 386 POTENTIAL.
FT CA_BIND 439 447 POTENTIAL.
FT SITE 1017 1021 GFPRK MOTIF.
FT DISULFID 94 103 BY SIMILARITY.
FT DISULFID 140 162 BY SIMILARITY.
FT DISULFID 185 197 BY SIMILARITY.
FT DISULFID 485 490 BY SIMILARITY.
FT DISULFID 496 550 BY SIMILARITY.
FT DISULFID 615 621 BY SIMILARITY.
FT DISULFID 694 702 BY SIMILARITY.
FT DISULFID 846 904 INTERCHAIN (BY SIMILARITY).
FT DISULFID 911 916 BY SIMILARITY.
FT CARBOHYD 86 86 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 107 107 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 265 265 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 500 500 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 511 511 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 605 605 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 656 656 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 697 697 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 841 841 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 857 857 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 935 935 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 969 969 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 1022 1066 TRYQMPKHYAVRIEEREYPPPGSTLPTKKHWTWSQTR
DQYV -> ARTRALYEAKRQAKMKSQSPSETERLDDY
(in isoform alpha-3A).
/FTId=VSP_002721.
SQ SEQUENCE 1066 AA; 118697 MW; 18F008E5CDF6B8CE CRC64;

Query Match 2.3%; Score 7; DB 1; Length 1066;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 RRPRLR 10

Db 524 RRPRLR 530

```

RESULT 50
YEGE_ECOLI
ID YEGE_ECOLI STANDARD; PRT; 1105 AA.
AC P38097; P76391;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yegE.
OS YEGE OR B2067.
GN Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saio N., Sampei G., Seki Y., Sivasubramanian S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RA "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392 (1996).
RN [3]
RP SEQUENCE OF 965-1105 FROM N.A.
RX MEDLINE=85054800; PubMed=6094528;
RA Nakabeppu Y., Miyata T., Kondo H., Iwanaga S., Sekiguchi M.;
RA "Structure and expression of the alkA gene of Escherichia coli
RT involved in adaptive response to alkylating agents.";
RL J. Biol. Chem. 259:13730-13736 (1984).
RN [4]
RP IDENTIFICATION
RX MEDLINE=95075659; PubMed=7984428;
RA Borodovsky M., Rudd K.E., Koonin E.V.;
RA "Intrinsic and extrinsic approaches for detecting genes in a
RT bacterial genome.";
RL Nucleic Acids Res. 22:4756-4767 (1994).
CC -!- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.
CC -!- SIMILARITY: Contains 3 PAS-associated C-terminal (PAC) domains.
CC -!- SIMILARITY: Contains 1 EAL domain.
CC -!- SIMILARITY: Contains 1 GGDEF domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AE000296; AAC75128.1; -
CC EMBL; D90844; BAA15920.1; -
CC EMBL; D90845; BAA15925.1; -
CC EMBL; K02498; -; NOT_ANNOTATED_CDS.
CC PIR; B64973; B64973
CC EcoGene; EG12396; yegE.
CC InterPro; IPR001633; EAL.
CC InterPro; IPR000160; GGDEF.
CC InterPro; IPR001610; PAC.
CC InterPro; IPR000700; PAS-assoc_C.

```

```

DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF00563; EAL; 1.
DR Pfam; PF00990; GGDEF; 1.
DR Pfam; PF05231; MASE1; 1.
DR Pfam; PF00785; PAC; 3.
DR Pfam; PF00989; PAS; 2.
DR SMART; SM00267; DUF1; 1.
DR SMART; SM00052; DUF2; 1.
DR SMART; SM00086; PAC; 3.
DR SMART; SM00091; PAS; 3.
DR TIGRFAMs; TIGR00254; GGDEF; 1.
DR TIGRFAMs; TIGR00229; sensory_box; 3.
DR PROSITE; PS50883; EAL; 1.
DR PROSITE; PS50887; GGDEF; 1.
DR PROSITE; PS50113; PAC; 3.
DR PROSITE; PS50112; PAS; 2.
KW Hypothetical protein; Repeat; Complete proteome.
FT DOMAIN 300 370 PAS 1.
FT DOMAIN 374 426 PAC 1.
FT DOMAIN 501 552 PAC 2.
FT DOMAIN 553 623 PAS 2.
FT DOMAIN 626 680 PAC 3.
FT DOMAIN 712 845 GGDEF.
FT DOMAIN 855 1104 EAL.
FT CONFLICT 965 966 EQ -> NS (IN REF. 3).
SQ SEQUENCE 1105 AA; 123886 MW; 22BE64B963CB9739 CRC64;

Query Match 2.3%; Score 7; DB 1; Length 1105;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 247 GGVLVVL 253
Db 139 GGVLVVL 145

Search completed: December 15, 2003, 14:59:29
Job time : 20 secs

```